Psychological Review

Memory Out of Context: Spacing Effects and Decontextualization in a Computational Model of the Medial Temporal Lobe

James Antony, Xiaonan L. Liu, Yicong Zheng, Charan Ranganath, and Randall C. O’Reilly

Online First Publication, July 25, 2024. https://dx.doi.org/10.1037/rev0000488

CITATION

Antony, J., Liu, X. L., Zheng, Y., Ranganath, C., & O’Reilly, R. C. (2024). Memory out of context: Spacing effects and decontextualization in a computational model of the medial temporal lobe.. *Psychological Review*. Advance online publication. https://dx.doi.org/10.1037/rev0000488

AMERICAN PSYCHOLOGICAL ASSOCIATION

Psychological Review

© 2024 American Psychological Association
ISSN: 0033-295X

https://doi.org/10.1037/rev0000488

Memory Out of Context: Spacing Effects and Decontextualization in a Computational Model of the Medial Temporal Lobe

James Antony[1, 2, 3], Xiaonan L. Liu[4], Yicong Zheng[2, 3], Charan Ranganath[2, 3], and Randall C. O'Reilly[2, 3]
[1] Department of Psychology and Child Development, California Polytechnic State University
[2] Department of Psychology, Center for Neuroscience, University of California Davis
[3] Department of Computer Science, Center for Neuroscience, University of California Davis
[4] Department of Psychology, The Chinese University of Hong Kong

Some neural representations gradually change across multiple timescales. Here we argue that modeling this "drift" could help explain the spacing effect (the long-term benefit of distributed learning), whereby differences between stored and current temporal context activity patterns produce greater error-driven learning. We trained a neurobiologically realistic model of the entorhinal cortex and hippocampus to learn paired associates alongside temporal context vectors that drifted between learning episodes and/or before final retention intervals. In line with spacing effects, greater drift led to better model recall after longer retention intervals. Dissecting model mechanisms revealed that greater drift increased error-driven learning, strengthened weights in slower drifting temporal context neurons (temporal abstraction), and improved direct cue–target associations (decontextualization). Intriguingly, these results suggest that decontextualization—generally ascribed only to the neocortex—can occur within the hippocampus itself. Altogether, our findings provide a mechanistic formalization for established learning concepts such as spacing effects and errors during learning.

Keywords: spacing effect, temporal context, error-driven learning, computational modeling, neurobiological drift

A primary goal in learning is to make information accessible long after encoding. One well-known technique for enhancing retention is distributing learning events over time rather than cramming them into a short interval. This phenomenon, known as the spacing effect (Maddox, 2016; Russo et al., 1998), is ubiquitous across many memory paradigms (Cepeda et al., 2006; Russo et al., 1998; C. D. Smith & Scarf, 2017) and operates over a wide spectrum of time scales spanning seconds (Glenberg, 1976), days (Cepeda et al., 2006; Küpper-Tetzel et al., 2014), months (Cepeda et al., 2008, 2009), and years (Bahrick et al., 1993).

While the spacing effect has profound real-world implications, it also presents a theoretical puzzle: How does the passage of time, which normally leads to forgetting, also allow for dramatically better long-term memory strengthening after further learning? We propose that one part of the answer is that forgetting can actually enable stronger learning the next time, to the extent that learning is based on the *difference* between the existing memory representation from prior learning and the representation at the time of relearning. This type of learning is known as error-driven learning (EDL), which plays a central role in our model. The other key element to this puzzle involves the role of *context* in both memory encoding and recall. Decades of research have shown that spatial, temporal, situational, and mental contexts contribute to the what–when–where of episodic memories for everyday learning events (Davachi, 2006; Eichenbaum et al., 2007). Each of these contextual factors can support holistic episodic memory recall when cued later, and they are each represented within the intricate neural machinery of the hippocampus (HC) and its major input, the entorhinal cortex (EC). However, just as contextual cues support memory, we will argue that they also limit the potential use of memories to the instances in which they can be reinstated. In order for learning to be accessible over longer intervals, memories may benefit from becoming temporally abstracted or "decontextualized"—two ways that the memory can generalize beyond the local, learned context (Karpicke et al., 2014; S. M. Smith & Handy, 2014). As explained further below, these processes constitute a major part of how our computational model produces spacing effects.

James Antony https://orcid.org/0000-0003-0656-2170

This article was funded by Office of Naval Research (Grant N00014-20-1-2578) awarded to Charan Ranganath. R. C. O'Reilly is Chief Scientist at the Astera Obelisk lab and eCortex, Inc., which may derive indirect benefit from the work presented here. The other authors declare no competing interests.

This work was previously presented at the following meetings: the 2022 Cognitive Neuroscience Society in San Francisco, CA; the 2022 Context Episodic Memory Symposium in Philadelphia, PA; and the 2023 International Conference on Learning & Memory in Huntington Beach, CA. It was also previously posted as a preprint at https://www.biorxiv.org/content/10.1101/2022.12.01.518703v2.

James Antony played a lead role in data curation, formal analysis, investigation, methodology, project administration, resources, visualization, writing–original draft, and writing–review and editing. Xiaonan L. Liu played a supporting role in formal analysis and methodology. Yicong Zheng played a supporting role in formal analysis and methodology. Charan Ranganath played a lead role in funding acquisition, a supporting role in conceptualization and writing–review and editing and an equal role in supervision. Randall C. O'Reilly played a supporting role in conceptualization, funding acquisition, and writing–review and editing and an equal role in supervision.

Correspondence concerning this article should be addressed to James Antony, Department of Psychology and Child Development, California Polytechnic State University, 1 Grand Avenue, San Luis Obispo, CA 93407, United States. Email: jwantony@calpoly.edu

Among the many dimensions of context, the most relevant for the spacing effect is temporal context. A collection of computational models have posited that, as memories are encoded, temporal context can be represented as a distributed, drifting pattern of neural activity (Balota et al., 1989; Estes, 1955; Horner et al., 2016; Howard & Kahana, 2002; Howard et al., 2008; Kahana, 2020; Kiliç et al., 2013; Lohnas et al., 2015; Mensink & Raaijmakers, 1988; Mozer et al., 2009; Murdock, 1997; Polyn et al., 2009; Raaijmakers, 2003; Rouhani et al., 2020; Sederberg et al., 2011). One consequence of this arrangement is that idiosyncratic, encoding-related activity patterns become reinstated during retrieval (El-Kalliny et al., 2019; Folkerts et al., 2018; Howard et al., 2012; Manning et al., 2011). In our model, this drifting temporal context provides a well-established explanation for the temporal forgetting function in terms of the gradually diminishing contextual cue support for the memory (Bouton, 1993; Crowder, 1976; Estes, 1955; Gershman & Niv, 2010; Mensink & Raaijmakers, 1988). Moreover, greater drift creates greater mismatches between the temporal contexts at encoding and relearning, enhancing the EDL that then drives greater plasticity for more widely spaced items. In addition, we show that this form of learning favors the elements in common between the two learning events, resulting in better subsequent recall that relies less on the temporal context. Thus, our model demonstrates how spacing effects emerge as a synergistic interaction between contextual drift and EDL (Mozer et al., 2009).

In the following introductory sections, we first discuss prior spacing effect findings and theories. Second, we provide behavioral and neural evidence that temporal context drifts across multiple time scales. Third, we tie these various lines of evidence together to introduce a drifting, biologically plausible model of the EC and HC to simulate paired associate learning. Finally, we discuss how and what gets strengthened due to spacing, how this compares against prior spacing effect theories, and what this means for the fate of the memory.

The Nonmonotonic Relationship Between Spacing and Final Retention Interval and Its Explanation

Spacing effects have received considerable attention in the cognitive psychology literature, generating a rich array of findings and explanations. Typically, longer intervals between learning events (interstimulus intervals, or ISIs) produce superior memory after some retention interval (RI) following the last instance of learning. However, one curious result is that more spacing is not always better. Rather, the optimal ISI depends strongly on RI, such that for very short RIs, short ISIs (or "massed" trials) often confer an advantage over spacing (Balota et al., 1989; Glenberg, 1976; Peterson et al., 1963; Rawson & Kintsch, 2005; Spieler & Balota, 1996; Toppino & Gerbier, 2014). In fact, plots relating ISI to RI are often nonmonotonic, with recall (given the same RI) rising sharply, reaching a maximum, and slowly decreasing with increasing ISI (Cepeda et al., 2006, 2009, 2008). No single ISI always benefits memory the most, and therefore, memory performance cannot be explained by a single factor—the relationship is more complex.

One prominent explanation for spacing effects is encoding variability theory. This theory suggests that a greater temporal difference between learning events results in more unique temporal elements assigned to each learning instance of the memory. This creates more variable encoding contexts during learning that, in turn, allow for more routes to the memory during retrieval (Balota et al., 2007; Glenberg, 1976, 1979; Huff & Bodner, 2014; Lohnas et al., 2011; McFarland et al., 1979; Melton, 1970; Raaijmakers, 2003; Ross & Landauer, 1978). However, as discussed above, maximum spacing (and hence, maximum encoding variability) does not always produce maximum memory benefits, making encoding variability alone unsatisfying as an explanation of spacing. Therefore, encoding variability has often been paired with another process called study-phase retrieval (Benjamin & Tullis, 2010; Cepeda et al., 2009; Greene, 1989; Mozer et al., 2009; Raaijmakers, 2003). Study-phase retrieval suggests that memory strengthening only occurs if elements of the study phase can actually be retrieved and reactivated during relearning (Thios & D'Agostino, 1976). Given that this ability will decrease over time, there will be a lower likelihood of strengthening at later intervals. On its own, study-phase retrieval would produce an *antispacing* effect, but combining it with the benefits afforded by encoding variability offers a plausible account of how one might observe nonmonotonic effects between ISI and RI in the spacing effect (Benjamin & Tullis, 2010; Landauer, 1969; Maddox, 2016; Raaijmakers, 2003).

While encoding variability is plausible in explaining many of the behavioral effects of spacing, the spacing effect is a temporal phenomenon and encoding variability largely does not incorporate recent developments on the critical role of temporal context for memory (though see Raaijmakers, 2003). Here, we will argue that the nonmonotonic nature of spacing effects can be explained via EDL as a strengthening mechanism. We propose that adding more routes to a memory trace, which is the memory-strengthening mechanism proposed by encoding variability theory, may be less important than strengthening aspects of the trace in common across learning events.

Behavioral and Neural Evidence of Multiscale Drift in Temporal Context

A central feature of episodic memory is that it is temporally organized. One classic example demonstrating this organization is that, during free recall, subjects tend to recall information successively that was presented nearby (Antony et al., 2021; Healey, 2018; Healey et al., 2019; Heusser et al., 2018; Howard & Kahana, 2002; Kahana, 1996, 2020; Uitvlugt & Healey, 2019). Explanations of this effect center around the temporal context model (Howard & Kahana, 2002; Lohnas et al., 2015; Polyn et al., 2009; Sederberg et al., 2008), which shows that a slowly drifting temporal context representation at encoding can become reinstated with a corresponding episodic memory, guiding subsequent memory recall to other information learned contiguously in time. Importantly, evidence for behavioral temporal contiguity spans numerous time scales. For instance, Howard et al. (2008) had subjects learn a number of word lists, and after the final list, they were asked to recall all words from all lists. They found that subjects did not only transition to nearby words during recall but also to nearby lists, suggesting temporal context has the property of being scale-invariant. Evidence for scale-invariant temporal representations has also arisen in behavioral paradigms like temporal estimation (Gibbon, 1977; Gibbon et al., 1984; Lewis & Miall, 2009; Merchant et al., 2013), other memory paradigms (Brown et al., 2007; Singh et al., 2017), and a variety of neural data (Bright et al., 2020; Cao et al., 2022; Folkerts et al., 2018; Guo et al., 2021; Jeunehomme & D'Argembeau, 2020; Manning et al., 2011; Nielson et al., 2015;

Rossi-Pool et al., 2019; Yaffe et al., 2014). Accordingly, spectral temporal representations have been incorporated into a variety of computational models of time (Brown et al., 2007; Grossberg & Schmajuk, 1989; Howard, 2018; Howard & Kahana, 2002; Jacques et al., 2022; Lewandowsky et al., 2012; Liu et al., 2019; Miall, 1989; Rolls & Mills, 2019; Tiganj et al., 2015), and the organization of time has been related to other laws of scale-invariance such as the Weber–Fechner law of perception (Arzy et al., 2009; Brietzke & Meyer, 2021; Cao et al., 2022; Dehaene, 2003). Intriguingly, scale-invariant temporal representations could also explain the shape of forgetting: if temporal context provides cue support for memories, drift over a spectrum of time scales would produce forgetting curves resembling human episodic memory data—that is, forgetting would proceed rapidly, and then more slowly, like a canonical forgetting curve (Ebbinghaus, 1885; D. C. Rubin & Wenzel, 1996).

Scale-invariant temporal context theories of memory have received support from recent neurobiological studies of the EC and HC. EC neurons drift, or slowly increase or decrease in activity, at varying rates over time, from seconds to hours (Aghajan et al., 2023; Bright et al., 2020; Tsao et al., 2018; Umbach et al., 2020). For example, Tsao et al. examined lateral EC (LEC) neurons in rats as they repeatedly explored two different environments over the course of an hour. Activity in a large proportion of recorded neurons drifted at varying rates, including some that drifted slowly over the entire session. Such an arrangement suggests the full population vector of LEC neurons drifts in a multiscale fashion over time. Within HC, representational drift of neuronal ensembles has been demonstrated over even wider scales, from seconds to months (Devalle & Roxin, 2022; Geva et al., 2023; Hainmueller & Bartos, 2018; Keinath et al., 2022; Kinsky et al., 2018; J. S. Lee et al., 2020; Liu et al., 2022; Mankin et al., 2015; Manns et al., 2007; Mau et al., 2018, 2020; A. Rubin et al., 2015; Umbach et al., 2022; Y. Ziv et al., 2013), and distinct memories acquired within short temporal windows share greater HC representational overlap (Cai et al., 2016; Rashid et al., 2016; Shen et al., 2022). Additionally, HC in rodents supports temporal order memory (Dusek & Eichenbaum, 1997; Fortin et al., 2002) and has cells that chart out repeated intervals of time, or “time cells” (Liu et al., 2019; MacDonald et al., 2011; Pastalkova et al., 2008; Reddy et al., 2021; Shimbo et al., 2021). Evidence from human functional magnetic resonance imaging (fMRI) experiments also suggests EC and HC support temporal representations. EC may support judgments of temporal duration (Lositsky et al., 2016), and anterolateral EC (the analogue of LEC that can be measured in human fMRI) may aid in recalling the temporal context of a movie (Montchal et al., 2019) and representing the temporal proximity of experience during learning (Bellmund et al., 2019). Additionally, HC in fMRI is sensitive to short temporal durations (Barnett et al., 2014), temporal proximity (Dimsdale-Zucker et al., 2022; Ezzyat & Davachi, 2014), sequences (Hsieh et al., 2014), and the time since encoding (Jenkins & Ranganath, 2010; Nielson et al., 2015). Altogether, these results suggest that the neural substrate for representing the temporal context of episodes, as simulated in computational models, may be instantiated in LEC and HC (see also Noulhiane et al., 2007).

Present Work: A Drifting Model of the EC and HC

The present modeling work investigates forgetting and spacing effects within a biologically realistic computational model of the medial temporal lobe, drawing inspiration from computational models of drifting, multiscale temporal contexts (e.g., Estes, 1955; Howard & Kahana, 2002; Mensink & Raaijmakers, 1988; Tiganj et al., 2015) and neurobiological evidence of drift (e.g., Tsao et al., 2018). Our model builds on prior complementary learning systems (CLS) models (Ketz et al., 2013; McClelland et al., 1995; K. A. Norman & O’Reilly, 2003; Rudy & O’Reilly, 2001) and includes neural network layers for HC subregions like the dentate gyrus (DG), cornu ammonis 3 (CA3), and CA1, which play distinct roles in episodic memory (Hasselmo & McGaughy, 2004; Leutgeb et al., 2004; Schapiro et al., 2017). The model also includes EC, which provides the main input into HC and receives its outputs (Witter et al., 2017), and we now include multiscale temporal context inputs to EC, simulating the timing of various learning schedules resembling human experiments. We therefore call it the HipSTeR (Hip-pocampus with Spectral Te-mporal Representations) model. The CLS configuration has multiple advantages in explaining episodic memory: It demonstrates how some episodic memory effects arise via specialized machinery, as in how high inhibition in the DG allows similar patterns to be separated (pattern separation; Grossberg, 1982; O’Reilly & McClelland, 1994; Rolls, 1989; Wigström, 1973) and how an area with strong intraconnections (like CA3) allows prior patterns to be recovered given incomplete inputs (pattern completion; Colgin et al., 2008; Hasselmo & McGaughy, 2004; Marr, 1971; K. A. Norman & O’Reilly, 2003; Rolls & Kesner, 2006; Treves & Rolls, 1994; Whittington et al., 2020); it shows how new learning can avoid catastrophic interference of old information by including contextual information as an input (Masse et al., 2018; O’Reilly & Munakata, 2000); and it elucidates how different subregions and pathways of the HC contribute independently to episodic memory effects (Schapiro et al., 2017; Zheng et al., 2022). Importantly, the model learns via a balance of associative (or Hebbian) learning and EDL processes (O’Reilly & Munakata, 2000), the latter of which we will show to be particularly relevant for simulating spacing effects.

Previous explanations of spacing effects center around encoding variability, whereas in our model, spacing effects arise largely via EDL. EDL conceptually dates back at least to learning rules created by Rescorla and Wagner (1972) and has been applied to learning & memory domains (Brod, 2021; Ku et al., 20211; Metcalfe, 2017) and computational models (Sutton & Barto, 1981; Zheng et al., 2022). Principally, in EDL, network weights that underlie memory traces change proportionally to the difference between network predictions based on activations and stored weights versus actual outcomes driven by current inputs (Ketz et al., 2013; O’Reilly, 1996; Zheng et al., 2022). As a memory-strengthening mechanism, encoding variability shares some similarities with the error-based account we posit in that it explains how greater benefits should accrue with greater spacing. However, the reasons for these benefits differ between the two accounts. Encoding variability predicts that spacing produces more contextual routes to retrieval. This largely assumes that the two events are independent at encoding rather than the later event at least partially updating the memory (Hintzman, 1986; Ross & Landauer, 1978). Note that the independence assumption is problematic due to the spacing property of superadditivity. Superadditivity refers to a higher probability of recalling a repeated item than either of two items separated by the same temporal lag (Begg & Green, 1988; Benjamin & Tullis, 2010; Maddox, 2016). That is, recall is above what would be obtained if you accounted for the recall likelihoods of two independent stimuli, indicating that there is no independence—rather, the original trace is reactivated. Conversely, the error-based account predicts first that the two learning instances interact (the later instance reactivates the first) and, more specifically, that greater temporal mismatch will produce stronger weight changes to the earlier trace via EDL. Critically, these greater weight changes will strengthen units in common across the memories. In other words, encoding variability predicts spacing effects arise via *more* connections, whereas EDL predicts they arise via *stronger* connections.

If temporal drift occurs, which elements of a prior memory remain in common during a relearning event? We discuss two strengthening processes relevant to this idea: temporal abstraction (Toppino & Gerbier, 2014) and decontextualization (S. M. Smith & Handy, 2014). Regarding temporal abstraction, if drift occurs over multiple time scales, greater spacing between training examples will result in relatively more overlap in slow-drifting units than in fast-drifting units. Although there will be a high temporal mismatch and, therefore, greater EDL, the weight changes from fast-drifting neurons will keep strengthening new units because they will have drifted too substantially to strengthen the old ones. However, because slow-drifting units retain greater overlap with the prior training examples, the greater EDL can actually strengthen these prior connections. As a result, greater spacing preferentially improves the weight strengths of units with longer time scales. Temporal abstraction fits nicely with rational accounts of spacing effects, which posit that, given computational constraints, it may be beneficial to support memory in proportion to how long it has been since it was last encountered (Anderson & Milson, 1989; Brea et al., 2014; Kording et al., 2007; Mozer et al., 2009). That is, information repeated numerous times within a short interval (e.g., hourly) and then not again will likely only be relevant for the next few hours, whereas information repeated on broader interval (e.g., monthly) will likely be important for longer, and it would be optimal to strengthen units according to this expectancy.

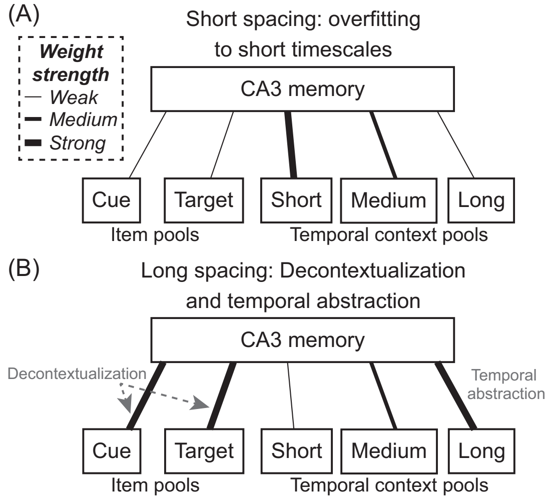
Decontextualization takes the idea of abstraction one step further, suggesting that memories may become resistant to all contextual drift. This process occurs via direct strengthening between the core elements of memory itself rather than a preferential strengthening of any manner of temporal context units; in the case of paired associate learning (which will be the focus of our modeling efforts), this involves direct weights between cue and target representations. Behavioral evidence for decontextualization from different kinds of contexts shows that, in comparison to constant learning contexts, variable learning contexts slow down learning but improve final memory performance when tests are given in a new context (Glenberg, 1979; Imundo et al., 2021; S. M. Smith et al., 1978; S. M. Smith & Handy, 2014, 2016). In other words, variable learning contexts make the memory depend less on elements of the original learning context at retrieval. Intriguingly, temporal spacing and environmental context change additively benefit memories tested days later in a novel context (S. M. Smith, 1982; S. M. Smith & Rothkopf, 1984). However, to our knowledge, decontextualization has not been used as a concept to explain spacing effects, whereby memories become less reliant on cue support from the temporal context for retrieval.

As a result of these strengthening processes, we suggest that "cramming" trials in time (relative to distributing them) produces *overfitting* to a local temporal context. These short ISIs can benefit memories when RIs are also short. However, after substantial temporal drift occurs (like with long RIs), the overfitting from short ISIs ultimately prevents the information from remaining accessible, whereas the temporal abstraction and decontextualization processes that occur with long ISIs keep the memory accessible (Figure 1). Therefore, we will argue that drift is not merely noise or a nuisance but may optimize memory function within a computationally constrained system. That is, drift allows old, nonrepeated information to be rationally forgotten while strengthening repeated information according to its temporal regularity (Anderson & Milson, 1989; Mozer et al., 2009).

Notably, the error-based model of spacing effects we propose here has an important and related antecedent. Mozer et al. (2009) simulated the spacing effect using a multiscale neural network model, whereby errors in representing temporal context at one timescale of drift were passed to representations at a longer timescale. Alongside the similarities, our approach also differs from theirs in numerous ways: (a) we simulated additional spacing effects; (b) we used a biologically realistic framework of EC–HC, which allowed us to more closely link the network's learning mechanisms to known neurobiology and create testable neural predictions; (c) temporal abstraction, which was built specifically in Mozer et al. (2009) by passing errors up one layer of the temporal hierarchy, emerged spontaneously in our model without such engineering; and (d) our model suggested decontextualization can also drive spacing benefits. In so doing, our model builds bridges between more abstract models of spacing effects (Mozer et al., 2009; Raaijmakers, 2003; Walsh et al., 2018) and the underlying neural mechanisms in the EC-HC, which creates new avenues for testing the underlying learning mechanisms and their implications for real-world memory performance.

Figure 1

Schematic of Basic Learning Principles Under Paired Associate Learning in Hippocampal Area CA3



Note. (A) Under low spacing, training will preferentially strengthen short timescale representations to CA3, effectively overfitting to the trained temporal context. (B) Under high spacing, greater error-driven learning will be allocated to associations between cues and targets to CA3 (decontextualization, stronger weights on left) and longer timescale representations (temporal abstraction, stronger weights on right). Both processes allow memories to remain accessible after greater drift occurs. CA3 = cornu ammonis 3.

Method

Model Architecture

Learning in neural networks occurs via the modification of synaptic weights between sending and receiving neurons. Our model was implemented in the Leabra (local, error-driven, and associative, biologically realistic algorithm) framework, which features two distinct learning rules. The first is Hebbian learning, which posits that changes to weights between connected units are incrementally updated through simultaneous, repeated activations (Hebb, 1949). The second, more powerful learning rule is EDL. This rule posits that the network constantly produces expectations (based on activations and stored weights) that are measured against outcomes and that learning is proportional to the difference between the two (O'Reilly & Munakata, 2000). The model also used rate-coded neurons separated into different layers and pools, sparse and distributed representations, competition driven by inhibitory interneurons within and across layers, and full bidirectional connectivity between some layers. Our specific model of the HC was based upon early CLS models (K. A. Norman & O'Reilly, 2003; O'Reilly & Rudy, 2001), with additions of theta-phase dynamics (Ketz et al., 2013) and EDL from the entorhinal cortex input layer (ECin) to CA3 (Zheng et al., 2022). The main changes in the present model are the expanded notion of temporal context into various spectra following evidence for this in EC (Bright et al., 2020; Tsao et al., 2018; Umbach et al., 2020) and the continuity of time across different learning epochs and tests. For this reason, we call this the HipSTeR (Hip-pocampus with Spectral Te-mporal Representations) model. Please see the Appendix for many of its parameters and https://github.com/CCNLab and https://github.com/JamesWardAntony/HipSTeR for all code and detailed explanations, including annotations, fully documented equations, and example simulations, including the model.

Temporal representations were divided into pools that shared a common underlying rate of drift, which involved random bit flips in activation/deactivation at each time step (Estes, 1955). In this way, time was translated into a spatial code that changed at each moment (Buonomano & Merzenich, 1995). To equalize the overall level of activation over time, whenever an active neuron became inactive, a random, previously inactive neuron became active. Following evidence that other EC cells (grid cells) have discrete spatial frequencies (Stensola et al., 2012; Wei et al., 2015; Whittington et al., 2020) and that drifting EC cells may have discrete drift rates (Aghajan et al., 2023), drift rates spanned a discrete set of values along a spectrum of timescales separated by powers of 2, with 1/4 (2^2) of the neurons in the fastest pools flipping at each time step and 1/512 (2^9) in the slowest drifting pools. Differences between the imposed drift rate and actual drift at each time step were carried over into successive time steps so that, especially in the slowest pools with drift rates of less than one unit per time step, drift would eventually occur at an approximately consistent rate across time. Figure 2A depicts the representational drift within each pool via its respective autocorrelation (Pearson r) against an initial timepoint ($t = 0$).

Drift was implemented both within lists across successive trials and between lists (Figure 2B). Within lists, successive trials always occurred after one time step, so the full list of 16 paired associates always spanned 16 time steps. Accounting for the idea that focused tasks or events create some stable state (Antony et al., 2021; Baldassano et al., 2017), after which larger "shifts" in activity occur (DuBrow et al., 2017), within-list drift was set to 1/4 the rate of between-list drift. Between-list drift occurred to varying extents depending on the experimental condition except in the No drift comparison condition.

The HipSTeR architecture (Figure 2C) involved the following layers:

1. An input layer comprising 16 pools of 49 neurons each: four cue pools, four target pools, and eight temporal context pools (Figure 2B). These pools were organized separately to reflect how information coming from various cortical regions converges in the HC (Eichenbaum et al., 2007). This layer has only direct, one-to-one forward connections with ECin. Additionally, this arrangement of separating temporal context from other inputs reflects the fact that multidimensional representations of global drift over time versus externally driven contextual factors (e.g., spatial environment) remain largely orthogonal (Keinath et al., 2022).

2. An ECin layer, which receives these signals from the input layer and sends connections into DG, CA3, and CA1 of the HC. ECin, which resembles superficial EC layers (Witter et al., 2017; Zhang et al., 2023), also receives inputs from ECout, which serves as the initial output layer from the HC and resembles deep layers of EC. ECin projections to DG and CA3 via the perforant path are broad and diffuse, with a 25% chance of connection.

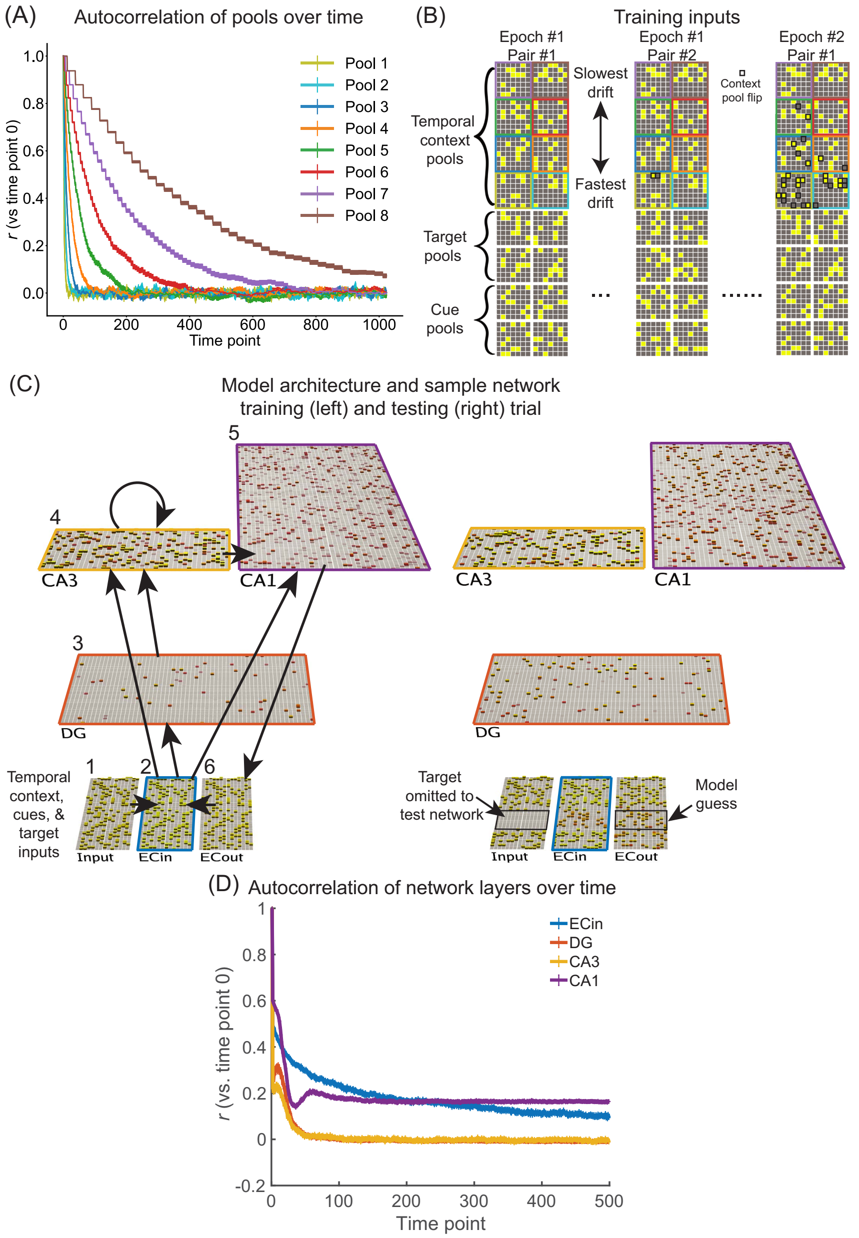
3. A large DG layer, which features high inhibition (Coulter & Carlson, 2007; O'Reilly & Munakata, 2000), resulting in only very sparse representations that functionally separate the patterns of similar inputs from ECin. In HipSTeR, DG has an inhibitory conductance multiplier of 2.9, resulting in the activity of about 1% of neurons. In turn, DG sends outputs to CA3 via strong mossy fiber projections (Henze et al., 2002; Vyleta et al., 2016), which have a multiplier of four and give it a stronger influence on CA3 activity than the more direct perforant path inputs from ECin.

4. A CA3 layer, which receives inputs from both ECin and DG and projects to itself (via recurrent collateral connections) and to CA1. The recurrent collaterals—which are fairly strong, with a strength multiplier of two in the model—have been theorized to be important for pattern completion because an activated representation can retrieve its previously learned association within this layer (Marr, 1971; O'Reilly & McClelland, 1994).

5. A CA1 layer, which receives and compares input from ECin and CA3, therefore serves as the convergence point for two hippocampal pathways and sends information back out of the HC to the entorhinal cortex output layer (ECout). The pathway from CA3, commonly referred to as the trisynaptic pathway (ECin → DG → CA3 → CA1 → ECout), essentially separates common inputs, binds items to contexts, and completes previously stored patterns based on degraded inputs. Evidence for rapid trisynaptic learning comes from findings such as the high inhibition featured in DG, which allows for a separation between

Figure 2

Temporal Representations and Their Implementations in HipSTeR



Note. (A) Temporal representations were separated into different pools of activity that simulated relatively independent cortical inputs and drifted at different rates. Autocorrelations of each temporal context pool were plotted over time against their initial time point. (B) Inputs to our complementary learning systems (CLS) model comprised the eight temporal context pools and four pools each representing cues and targets for paired associate learning. Shown are two successive cue–target pairs during a first training epoch and the first pair again in a second training epoch. Black squares indicate flips in the context pools, showing how drift proceeds differently across pools. (C) Training inputs from (B) entered HipSTeR via ECin, after which they entered the hippocampal loop, with all model connections depicted by arrows on the left. During training (left), the model learned the input patterns. During testing (right), it was presented with the cue and current temporal context pools but without the target pools. Performance was measured by how well the model reproduced target pool activity in ECout. Numbers correspond to layer numbers in the detailed explanations within the Methods–Model Architecture section. (D) Autocorrelations for the primary network layers are depicted in (A) as the correlation of the activation pattern in each layer against its pattern on the first time step. This was run on a separate, "pure" version of the model without any cue–target repetitions, meaning there were 500 unique cue–target pairs alongside identical temporal context pools. Colors correspond to those surrounding each network layer in (C). HipSTeR = Hip-pocampus with Spectral Temporal Representations; DG = dentate gyrus; ECin = entorhinal cortex input layer; ECout = entorhinal cortex output layer; CA3 = cornu ammonis 3; CA1 = cornu ammonis 1. See the online article for the color version of this figure.

highly similar pattern inputs (Leutgeb & Leutgeb, 2007; Vazdarjanova & Guzowski, 2004), the importance of area CA3 in learning new paired associates (Rajji et al., 2006), and the role of these regions together in discriminating between highly similar information in memory (Bakker et al., 2008). Note that here we suggest an important role for a disynaptic, ECin → CA3 → CA1 → ECout subpathway, following modeling results that this pathway can support generalization (Kang & Toyoizumi, 2024; Kowadlo et al., 2019) and learn via EDL (Zheng et al., 2022). The pathway from ECin → CA1 → ECout constitutes the monosynaptic pathway of the HC (Schapiro et al., 2017), which allows CA1 to directly encode target ECin activity (Grienberger et al., 2022) and sends activity from the HC back into the cortex. Evidence for slower monosynaptic learning comes from its having a slightly slower learning rate (Nakashiba et al., 2008) and more overlapping, generalized representations (Fenton et al., 2008; Leutgeb et al., 2004; Schapiro et al., 2017; Singer et al., 2010). These connections remain within pools, following their point-to-point anatomical connectivity patterns (Witter et al., 2017). This pathway, therefore, serves an autoencoder function, which translates the pattern-separated representations from the trisynaptic pathway back into a common reference frame for the cortex.

6. An ECout layer, which serves as the output of the HC and, therefore, the hippocampal network's "guess" during testing (Figure 2C). Additionally, it also serves as the input back into ECin, which can result in different activations in successive cycles through the HC (Kumaran & McClelland, 2012; Schapiro et al., 2017).

The effects of drift differed across layers of the network. We depicted this without any cue–target pair repetitions by training the network on different cue–target associations for each of 500 time steps and calculated the autocorrelation of each region with its initial time step. Of the hippocampal areas, CA1 showed the slowest pace of drift, followed by DG, which was followed very closely by CA3 (Figure 2D).

Model training and testing followed four discrete phases resembling activity during the four quarters of the hippocampal theta rhythm (Ketz et al., 2013). The model was learned via two EDL mechanisms. In the first mechanism, the first three quarters constituted what are considered the minus phases, whereby the network produced an expected output based on its weights and input activations. The fourth and final quarter was the plus phase, whereby the target activation was provided from ECin → ECout, and thereby, learning occurred based on the difference between the network's prediction from the minus phases into ECout and the actual outcome. The first and fourth theta phases came during theta troughs when CA1 was strongly influenced by ECin (Siegle & Wilson, 2014). Conversely, at theta peaks, CA1 was strongly influenced by CA3, which involved a guess based on activations and previously stored patterns. During the plus phase, ECin drove both CA1 and ECout activity, effectively clamping the correct answer in both EC layers and forcing weight adjustments in CA1. Therefore, across learning, ECout activity came to resemble ECin activity via the CA1 projection during the minus phases (without the direct ECin → ECout input). The second mechanism involved EDL in CA3 (Zheng et al., 2022). This error arose as a form of temporal difference learning between different pathways converging on CA3 neurons (Sutton & Barto, 1998): direct input from ECin (via the perforant path) and CA3 recurrent collateral activations arrived on CA3 neurons within the first quarter of the theta cycle, and critically, this input preceded signals from the multisynaptic ECin → DG → CA3 pathway (Yeckel & Berger, 1990). This minus phase constituted CA3 activity prior to DG inputs, and the plus phase occurred when they arrived. Therefore, the pattern-separated DG activation acted as a teaching signal to correct the predicted pattern in CA3 based only on perforant path and recurrent collateral activations (Kowadlo et al., 2019).

In our simulations, as in prior models (Ketz et al., 2013; Zheng et al., 2022), temporal context drift occurred within trials of an epoch. However, drift differed in HipSTeR in multiple ways. As mentioned above, drift occurred in all simulations across a spectrum of time constants in a manner that was constant within each temporal context pool. Additionally, we differentially modified drift in some experimental conditions in two other ways. First, drift often occurs between learning epochs, with the number of drifting time steps defined as those coming after the final learning trial of an epoch and before the first trial of the next epoch. We refer to this drift as the ISI. The exception to this was the No Drift condition, in which drift still occurred within-epoch, but each epoch of training was identical. Given that neurobiological drift occurs (e.g., Tsao et al., 2018), such a condition is biologically impossible, as an agent would never return to the exact same drifting neural pattern during relearning. Notably, similar neural firing patterns can recur even on long timescales with repeated experiences (Liu et al., 2022; Sun et al., 2020), but they are not identical. However, it is common to train neural networks this way, including in prior versions of the CLS architecture, so we used the No drift condition as an illustrative comparison.

Second, drift often occurred after the final training epoch and before the model was tested. We refer to this as the RI, following the labeling convention used in human behavioral experiments employing memory tests after various intervals. We defined the number of drifting time steps as those occurring after the final learning trial of the final training epoch and before the first trial of testing. Drift continued at the same rate for each trial of the testing epoch. Similar to the No drift training condition, we also had a No Lag (RI-0) RI, for which the temporal context given at the test was the exact temporal context used in the final training epoch. Similar to the No drift condition above, the No Lag condition is biologically impossible but likewise served as a useful control to assess the network's ability to recall under the exact conditions of at least one of its learning epochs.

HipSTeR has trisynaptic (ECin → DG → CA3 → CA1 → ECout), disynaptic (ECin → CA3 → CA1 → ECout) and monosynaptic (ECin → CA1 → ECout) pathways. Thus, it is possible to turn off learning (prevent all weight changes) in some pathways and reasonably expect some alternate learning to proceed (Schapiro et al., 2017). Additionally, ECin → CA3 learns via both Hebbian learning and EDL, which changes weights proportionally to the difference between CA3 activity with ECin, DG, and recurrent CA3 inputs present against activity prior to DG input. We can, therefore, isolate the importance of these learning rules within ECin → CA3 by turning off EDL while leaving Hebbian learning intact or turning off all learning. Altogether, we compared the full HipSTeR model to alternative models in which learning pathways were affected in the following ways: ECin → DG (no learning), ECin → CA3 (no learning), ECin → CA3 (no EDL, but Hebbian learning present), ECin → CA1 (no learning), CA3 → CA3 (no learning), and CA3 → CA1 (no learning). To assess the impact of multiscale drift versus other uniform drift implementations, we also compared simulations of our control HipSTeR model against networks wherein all temporal context pools drifted at a uniform slow, medium, or fast rate. The slow-drifting network used the slowest drift rate from HipSTeR (1/512 per time step), the fast-drifting network used the fastest drift rate (1/4), and the medium-drifting network used a medium rate (1/64).

Experimental Conditions in Spacing Effect Simulations

We will now outline specific simulations of prior behavioral findings. For these and other simulations, our hypotheses were not preregistered. More information about the original studies can be found in the corresponding region in the Results section. To model spacing effects with one variable ISI (Cepeda et al., 2008), we used four training epochs with a unique scheduling procedure. The first two epochs simulated the encoding and one perfect recall trial of the first learning session. These were implemented in direct temporal succession (ISI = 2 between lists) in the model. The third and fourth epochs simulated the two practice + feedback trials of the second learning session, and these also occurred in direct temporal succession (ISI = 2). Critically, the ISI between the second and third lists differed across experiments. These eight ISIs spanned powers of 2, from 4 (2^2) to 512 (2^9). We then used four RIs after the final training epoch spanning powers of 2, from 64 (2^6) to 512 (2^9). We chose these ISI and RI combinations so that we had a mix of ISI:RI ratios, ranging from far less to far greater than 1. We fit these data to find optimal ISIs using an equation from (Cepeda et al., 2008): $y = -a \times (log(x + 1) - b)^2 + c$, where y is recall performance, x is ISI; and a, b, and c are free parameters. From the best-fit parameters, we found the timepoint corresponding to maximum performance in the curve to obtain the optimal ISI.

For the spacing override effect (Rawson et al., 2018), three experimental groups had three epochs of training with lags of 2 (low spacing), 8 (medium spacing), or 32 (high spacing). Following this initial training session were two training epochs separated by 128 time steps each. After these epochs, the final test occurred after another 128 time steps. Note that the low and high spacing conditions approximately map onto Rawson et al. (2018) conditions of Lag-15 and Lag-47 because they used different naming conventions; by their conventions, our Massed would be Lag-17 (15 drifting time steps during the training list itself +2 time steps after the list), and our Lag-32 would be Lag-47 (15 during training + 32 after the list).

To model the importance of absolute amounts of spacing and different spacing regimens, including expanding, contracting, and equal spacing (Küpper-Tetzel et al., 2014), we used four different experiments, each using five training epochs with unique ISIs. All experiments had very short intervals between the first, second, and third training epochs. Following this, the first three experiments had intervals between training epochs that were expanding (16, 256), contracting (256, 16), or equally spaced intervals (136, 136) that matched the overall drift of the prior two (equal, match). To demonstrate the importance of absolute spacing, the final experiment (equal, compressed) used equally spaced intervals between training epochs but 1/8 of the amount of overall spacing (17, 17). After the final training epoch, RIs occurred after drift corresponding to values spanning powers of 2, from 32 (2^5) to 2048 (2^{11}) time steps.

To model list-wise spacing with repeated variable ISIs (Bahrick et al., 1993), we used five training epochs using consistent ISIs in each experimental condition. The drifting conditions had ISIs of powers of 2, including 8 (2^3), 64 (2^6), and 512 time steps (2^9). After the final training epoch, RIs occurred after drift corresponding to values of powers of 2, including 64 (2^6) and 1,024 (2^{10}). Additionally, a scrambled condition used scrambled temporal context vector pools. These pools were only scrambled once after training, meaning that they used the same drift rate during the testing epoch, which controlled for the presence of drift during testing.

In later simulations aimed at demonstrating the mechanisms of learning in HipSTeR, we primarily used the approach from modeling Bahrick et al. (1993; repeated variable ISI) with a few additions. First, we added more ISIs and RIs. ISIs spanned powers of 2, from 2 to 512 (2^9) time steps, while RIs went from 32 (2^5) to 2048 (2^{11}) time steps. Second, we added a No drift ISI condition that involved training the model using the same temporal context vectors in each epoch. Drift still occurred within this list, but the input pattern, for example, the first cue–target pair, was the exact same across epochs. Third, we added a Scrambled ISI condition, whereby we completely scrambled the temporal contexts before each new training epoch while preserving within-list drift within a training epoch. Finally, we added a No Lag RI condition that was tested using the temporal context vector from the final training epoch. We still used the Scrambled RI condition as in Bahrick et al. (1993) simulations that involved scrambling the temporal contexts before the final test.

Measuring Model Memory Performance, CA3 Error, Representational Similarity, and Weights

In each experimental condition, model weights were reinitialized and trained anew with random weights within what we call runs. Different runs were analogous to random assignment for human participants, and we, therefore, performed inferential statistics across different runs of the model.

We tested model memory performance by assessing the activation of ECout neurons after the minus phases against the plus phase. Neuronal activity above and below 0.5 was considered active and inactive, respectively. For correct performance on a given trial, the proportion of active neurons that were expected to be inactive and the proportion of inactive neurons that were expected to be active had to both be below 0.2.

To assess error across training epochs, we calculated the absolute difference in CA3 activation in the first quarter (Q1) versus the final plus phase (Q4) and divided this quantity by the average trial-wise CA3 activation. This error metric indicated how easily the network produced the intended output. Given that it generally scales with poor current performance but greater subsequent learning, it provides a useful proxy for the concept of desirable difficulties in learning, in which making learning more difficult often has positive long-term consequences for memory retention (Bjork & Bjork, 1992).

To gain a sense of how representations changed across epochs, we calculated the representational similarity in CA3 during all training trials. To do this, we correlated across-CA3 neuron activation patterns at the end of each trial in the current epoch (starting in Epoch 2) against each trial in the prior epoch. These values were then separated based on whether they involved the same versus different input cues.

To assess the structural changes to the network across learning and as a function of experimental conditions, we measured the average weight strength between ECin → CA3 neurons for each pool. We measured changes in this pathway (the perforant path) because of its role in supporting cue–target learning and its EDL properties. To assess how various experimental conditions affected representations in different temporal pools, we calculated the average strength from neurons in each ECin temporal pool separately. To assess the effects of training regimens, we used values after the final training epoch and contrasted various experimental conditions (e.g., Drift vs. No drift) across runs.

Experimental Conditions in Item-Wise and Block-Wise Decontextualization Simulations

Following evidence of decontextualization in the spacing effect simulations, we simulated more canonical decontextualization paradigms using epoch-wise (Imundo et al., 2021; S. M. Smith et al., 1978) or item-wise contexts (S. M. Smith & Handy, 2014, 2016). Rather than eight temporal context pools, our inputs here involved six temporal context pools and two other context pools representing either epoch-wise or item-wise context. For epoch-wise contexts, which were analogous to learning associations for an entire session within a spatial context (Imundo et al., 2021; S. M. Smith et al., 1978), the other context pools were either the same for all training epochs (constant epoch condition) or new for each epoch (variable epoch condition). For trial-wise contexts, which were analogous to having some incidental background context present during each individual learning association (S. M. Smith & Handy, 2014, 2016), the other context pools were unique for each cue–target pair but were either the same across each training instance of that pair (constant trial) or changed each time (variable trial). Finally, the other context pools at the test could either be the same as the context from the first training epoch (old test) or a new context (novel test). All context vectors were randomly generated and bore no resemblance to others. For these experiments, we used a short ISI (two steps) and moderate RI (512 steps) for all simulations to control for the temporal context.

Results

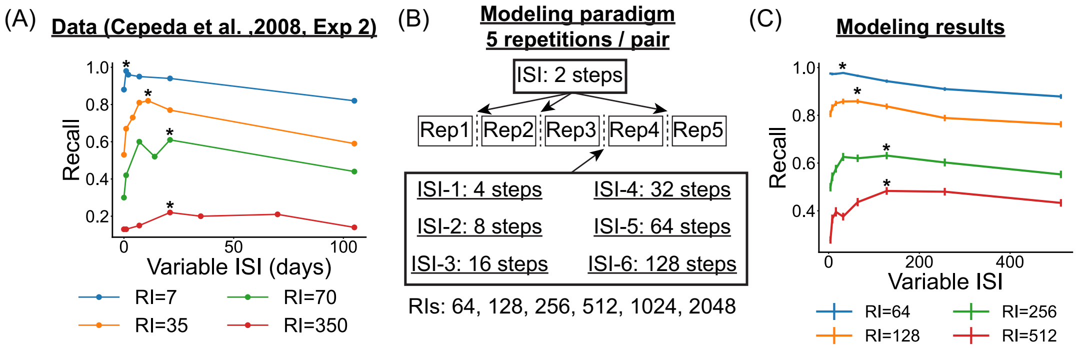
We began our investigations by simulating a number of behavioral findings from the spacing effect literature. Following this, we probed the mechanisms by which our HipSTeR model learned amid constantly drifting inputs during training, including assessments of error, representational similarity within layers of the network, weight changes, eliminating learning in specific HC pathways under various training regimens, and comparing our multiscale drift model to alternative models with uniform drift. Last, to connect our findings to more canonical ideas of decontextualization, we simulated decontextualization paradigms that were unrelated to temporal context.

Optimal ISI Decreases With RI and Optimal ISI:RI Ratio Decreases With Increasing RI: List-Wise Spacing With One Variable ISI

We first simulated an experiment (Cepeda et al., 2008) containing a single variable ISI separating two training sessions. Briefly, Cepeda et al. (2008) had human subjects learn lists of paired associates to a criterion of two correct trials and wait a variable number of days before performing a criterion of two more correct trials (0 [3 min], 1, 2, 7, 21, or 105 days), with testing occurring after RIs of another 7, 35, 70, or 350 days. They found that the relationship between ISI and memory recall was nonmonotonic (e.g., optimal at some ISI between the extremes of those tested) and that the optimal ISI increased with RI (Figure 3A). Additionally, they found an intriguing relationship between ISI and RI, such that the optimal ISI:RI ratio was not consistent but actually decreased with increasing ISI.

Figure 3

Simulations Showing Optimal ISI Depends on RI and Optimal ISI:RI Ratio Decreases Over Time



Note. (A) Data replotted from "Spacing Effects in Learning," by N. J. Cepeda, E. Vul, D. Rohrer, J. T. Wixted, and H. Pashler, 2008, *Psychological Science*, *19*(11), p. 1098 (https://doi.org/10.1111/j.1467-9280.2008.02209.x). (B) Our modeling paradigm, wherein only a single ISI between repetitions 3 and 4 differed across the conditions. Note that we used steps as a proxy for time in the behavioral experiments. (C) Modeling results. Data from simulations were plotted as mean ± SEM across runs in the model. In (A) and (C), the optimal ISI for the same RI is marked with an asterisk. Exp 2 = experiment 2; ISI = interstimulus interval; RI = retention interval; Rep = repetition; SEM = standard error measurement. See the online article for the color version of this figure.

We simulated these results by (a) training the model to learn lists in short succession (to capture initial study and one round of learning to criterion), (b) imposing one variable ISI (from 4 [2^2] to 128 [2^7] trials) between the second and third training epochs, (c) training the model on two more lists occurring within short succession (to capture the final rounds of learning), and (d) testing the model after variable RIs (from 64 [2^6] to 512 [2^9] steps; Figure 3B). First, we performed a 2-way, ISI × RI between-subjects analysis of variance (ANOVA), and we found a main effect of RI, $F(3, 3564) = 2962.0$, $p < .001$, a main effect of ISI, $F(8, 3564) = 20.5$, $p = .19$, and a significant interaction, $F(24, 3564) = 67.1$, $p < .001$. Second, similar to Cepeda et al. (2008), we found that the relationship between ISI and RI was nonmonotonic, peaking at medium (not the shortest nor the longest) ISIs for each RI. Third, the optimal ISI increased with increasing RIs; to find this, we used model fits based on a three-parameter equation from Cepeda et al. (2008), $y = -a \times (log(x + 1) - b)^2 + c$ (see the Method section for details). The optimal ISIs from these model fits were 14.3, 34.8, 70.7, and 204.1, respectively. Third, the ISI:RI ratio with increasing RI decreased, as these ratios were 64/14.3 = 4.48, 128/34.8 = 3.68, 256/70.7 = 3.62, and 512/204.1 = 2.5, respectively (Figure 3C). Therefore, we captured the main principles of the spacing effect outlined in Cepeda et al. (2008). Of these principles, the first is especially important conceptually because it suggests there is not one solo factor underlying memory strength—if there were, the ISI condition resulting in the greatest strengthening would produce the best memory performance regardless of RI. We will later demonstrate that these differentiating factors depend on the amount of overlap between the temporal context at test and the learned contexts (as governed by the RI), the strength of each of the weights in each of the layers, and the direct strength between the cue and target pools.

Relearning Override Occurs With Relatively Large Amounts of Spacing

If a memory has been learned with low spacing, can it still benefit from spacing later? Next, we addressed this question, following what has previously been referred to as the relearning override effect (Rawson & Dunlosky, 2011; Rawson et al., 2018). Briefly, this effect occurs when an initial, small difference in either spacing (Rawson et al., 2018) or initial learning (Rawson & Dunlosky, 2011; Rawson et al., 2018) becomes largely (but not necessarily completely) overridden by relearning after a longer spacing interval. That is, the relative gain after a longer spacing interval is larger for an initially weaker memory, whether it be weaker because of fewer initial learning trials or more massed training. To demonstrate the override effect in terms of initial differences in spacing, we turned to the learning criterion = three conditions of Experiment 1 in Rawson et al. (2018). In this condition and experiment, subjects initially studied Lithuanian–English word pairs with different initial lags of either 15 or 47 during the first session. They were then practiced in the same order until they were retrieved correctly thrice. After this session, subjects returned for a series of relearning sessions spaced 1 week apart (Figure 4A, right). The relative gain when relearning after a large temporal gap was larger for initially less-spaced memories.

To model the spacing override effect, we created three experimental conditions (Figure 4B). These conditions used three initial epochs of training with lags of 17 (15 for training list + two extra time steps; low spacing), 23 (15 for training list + eight extra time steps; medium spacing), and 47 (15 for training list + 32 extra time steps; high spacing). In all conditions, initial training was followed by two later training epochs, each after 128 time steps, followed by a final test after another 128 time steps.

Investigating memory at each training epoch in the first two groups, we first found a typical effect of faster learning in the lower spacing conditions before the first larger gap; final training epoch, $F(2, 297) = 194.2$, $p < .001$; Figure 4C. After this gap, memory was superior for the higher spacing conditions, $F(2, 297) = 122.6$, $p < .001$, in line with classic spacing effects. However, after two large gaps, memory benefits were higher for the initially low spacing groups at the final test, $F(2, 297) = 9.0$, $p < .001$; follow-up *t* tests for the low versus and medium and high spacing groups indicated both $t > 3$, $p < .003$, demonstrating a relearning override of initial, relatively small spacing.

In some ways, the override effect of initial learning criteria is surprising, as it has long been known that increasing the number of learning trials slows the rate of forgetting (e.g., Krueger, 1929). However, these benefits tend to be weak and transitory when overlearning occurs within close temporal succession (or all within the same session; Driskell et al., 1992; Elliott et al., 2014; Pyc & Rawson, 2009). Therefore, the relative change in memory change offered by later spacing is substantial enough to drastically reduce or eliminate these initial differences. These findings underscore the idea that learning that is relatively compressed in time *overfits* one temporal context that ultimately proves unhelpful when that temporal context is no longer active. It also points to the importance of absolute spacing, or the total time between the first and last training instances for a given memory, which we will cover in more detail in the following section.

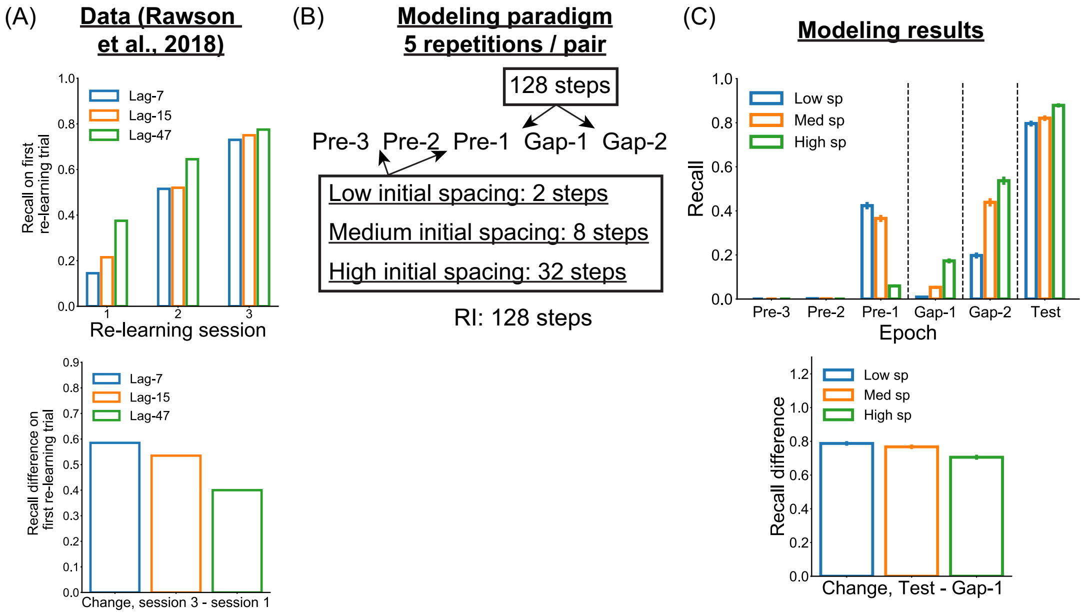
Alternative Schedules and Absolute Spacing: Contracting, Expanding, and Equally Spaced Intervals

The relearning override effects of large spacing suggest that the greatest temporal determinant of later memory may be the absolute amount of spacing between all training instances (Karpicke & Bauernschmidt, 2011). However, a number of investigations have examined the importance of alternative training schedules, such as ISIs that expand, contract, or remain equal across training instances (Gerbier et al., 2015; Mettler et al., 2016; Toppino et al., 2018; Toppino & Gerbier, 2014). In one study (Küpper-Tetzel et al., 2014), subjects learned paired associates before relearning on multiple sessions after either expanding (1-, 5-, and 9-day), contracting (9-, 5-, and 1-day), or equal (all 5-day) ISIs. Final tests were given either 1, 7, or 35 days after the final learning session. Free recall performance (requiring both items of a pair to be recalled and matched) showed that the contracting schedule was superior to the equal and expanding schedules for the early (1- and 7-day sessions) RIs, whereas equal and expanding schedules were superior to the contracting schedule at long RIs (Figure 5A; They focused on free recall because cued recall was at or near ceiling performance, though we will model cued recall).

We modeled the effects of alternative schedules and absolute spacing using four experimental conditions with unique ISIs (Figure 5B). Here the intervals before the two final training epochs differed in the following conditions by the number of drifting time steps: expanding (16, 256), contracting (256, 16), equal matched (136, 136), and equal compressed (17, 17). The first three conditions

Figure 4

Simulations Showing That Relearning Override Occurs When Later Spacing Is Greater Than Initial Spacing



Note. (A; top) Data replotted from "Investigating and Explaining the Effects of Successive Relearning on Long-Term Retention," by K. A. Rawson, K. E. Vaugh, M. Walsh, and J. Dunlosky, 2018, *Journal of Experimental Psychology*, *24*(1), p. 61 (https://doi.org/10.1037/xap0000146). (bottom) Recall differences from the first to third relearning sessions showed larger benefits in initially lower spacing conditions. (B) Our modeling paradigm, wherein we created low, medium, and high initial spacing conditions before longer gaps, relearning epochs, and a final test. (C) Modeling results. (top) Data from simulations were plotted as mean ± SEM across runs in the model, including performance on initial training runs (left), after larger gaps (middle), and at the final test (right; in some cases, error bars are too small to visualize). (bottom) Model recall differences from just after the first large gap (Gap-1) to the final test showed larger benefits with initially lower spacing. RI = retention interval; Pre = pregap learning repetition; sp = spacing; SEM = standard error measurement. See the online article for the color version of this figure.

allowed us to assess the importance of alternative schedules, while the final condition allowed us to again assess the importance of absolute spacing. Finally, RIs occurred after 32 (2^5) to 2048 (2^{11}) time steps.

Overall, a two-way learning schedule (expanding, contracting, equal, matched, or equal, compressed) × RI (32–2048) ANOVA revealed significant main effects of schedule, $F(3, 2376) = 21.7, p < .001$, RI; $F(5, 2376) = 671.2$, $p < .001$, and an interaction, $F(15, 2376) = 16.8$, $p < .001$. Similar to Küpper-Tetzel et al. (2014), the Contracting schedule demonstrated superior performance to the other conditions at the early RI: the order, considering significant differences at $p < .05$ as ">" and insignificant ones as "=", was equal, compressed > contracting > equal, matched > expanding (Figure 5C). We believe that this occurred because this group had the most training opportunities within a short temporal interval of these tests, allowing for better training within this narrow temporal context. However, at the longest RIs, we found that this advantage had disappeared, reversing against the equal condition and showing no difference from the expanding condition: The order was equal, matched > contracting = expanding ($p = .87$) > equal, compressed. Note that the insignificant differences between expanding and contracting at long RIs differ from the Küpper-Tetzel et al. (2014) results for their longest RI. All three groups showed superior memory at the longest RIs against the equal compressed group (all $p < .023$), once again highlighting the critical importance of absolute spacing when tests occur after long RIs.

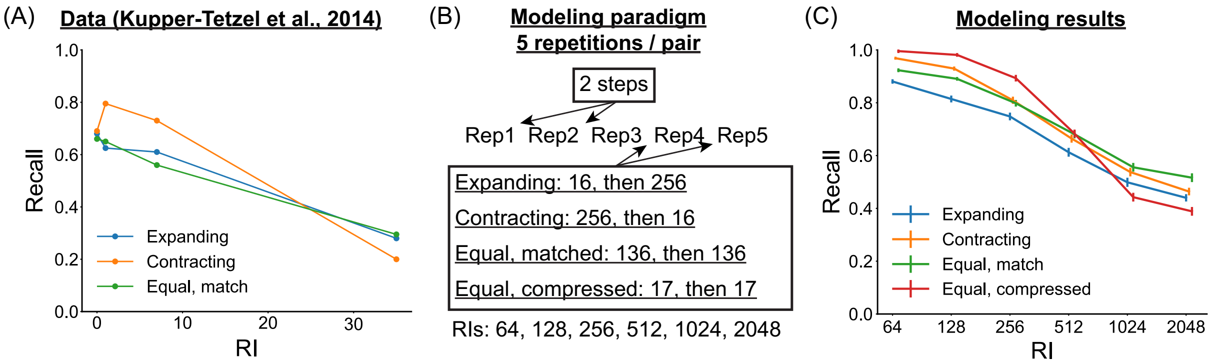
Therefore, we showed that contracting schedules had superior performance at short RIs, likely due to their having more training examples within close temporal proximity of these tests, but this advantage disappeared over time. Additionally, the equal spacing condition produced superior recall at some long RIs. One aspect of the results of Küpper-Tetzel et al. (2014) that we were not able to capture was superior memory for expanding than contracting schedules at the longest time points. Benefits for expanding over contracting schedules are not always found at the longest RIs (e.g., Cull, 2000; Karpicke & Bauernschmidt, 2011). However, we believe there is a larger point: the subtle differences in recall across RIs between these three conditions—as well as findings from the relearning override effects—point to the *absolute* spacing of training instances as the greatest determinant of long-term memory performance. This accords with behavioral findings showing that differences between expanding, equal, and contracting schedules were inconsistent and minor in comparison to differences at three levels of absolute spacing (Karpicke & Bauernschmidt, 2011).

Spacing Effects at Extremely Long RIs Using Repeated Variable ISIs

Another seminal spacing effect finding involved multiple learning episodes spread over ISIs of up to almost 2 months and final RIs of up to 5 years (Bahrick et al., 1993). In the study, four members of the same (Bahrick) family learned foreign language–English pairs

Figure 5

Simulations Showing the Importance of Absolute Spacing and Investigations of Alternative Schedules



Note. (A) Data replotted from "Contracting, Equal, and Expanding Learning Schedules: The Optimal Distribution of Learning Sessions Depends on Retention Interval," by C. E. Küpper-Tetzel, I. V. Kapler, and M. Wiseheart, 2014, *Memory and Cognition*, *42*(5), pp. 729–741 (https://doi.org/10.3758/s13421-014-0394-1). (B) Our modeling paradigm, wherein we created expanding, contracting, equal, and matched, or equal and compressed conditions to show the impact of different relearning schedules and absolute spacing. (C) Modeling results. Data from simulations were plotted as mean ± SEM across runs in the model (in some cases, error bars are too small to visualize). RI = retention interval; Rep = repetition; SEM = standard error measurement. See the online article for the color version of this figure.

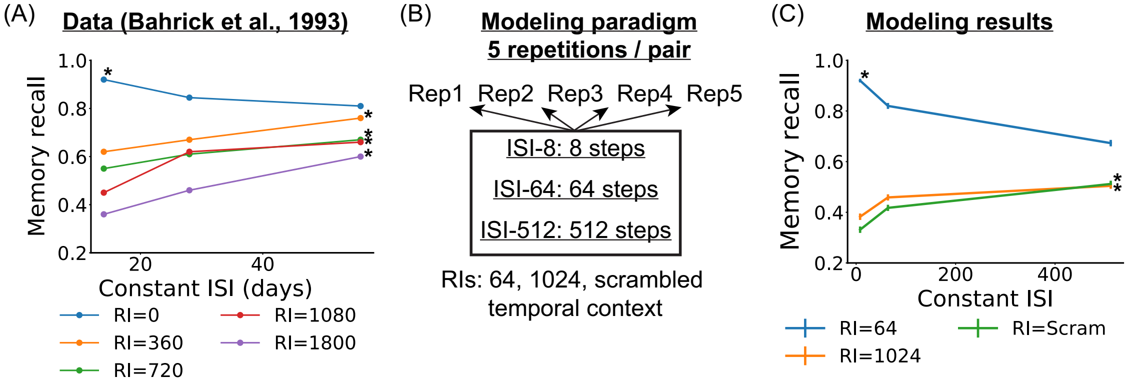
over 13 or 26 sessions spaced by 14, 28, or 56 days before a final test 1, 2, 3, or 5 years after the final learning session. Similar to prior effects showing that short ISIs improve memory at short RIs, performance was best at the end of training for the 14-day interval schedule. However, when assessed after RIs from 1 to 5 years, performance was best for words in the 56-, then 28-, then 14-day interval (Figure 6A). For our purposes, these findings show that spacing effects can compound over numerous intervals and can still be demonstrated at extremely long RIs.

We modeled these findings by training HipSTeR using the same ISI (8, 64, or 512 time steps) across each of five epochs (Figure 6B). After the final training epoch, we implemented RIs of 64 and 1,024 time steps. We additionally added a condition using scrambled temporal context vector pools to simulate what could arguably occur to temporal context at very long RIs (e.g., 5 years in Bahrick et al., 1993).

A two-way, ISI (8, 64, or 512 time steps) × RI (64, 1,024, or scrambled) ANOVA revealed significant main effects of ISI; $F(3, 888) = 137.7$; $p < .001$, RI; $F(2, 888) = 1233.5$; $p < .001$, and an interaction, $F(6, 888) = 109.3$; $p < .001$. Similar to Bahrick et al. (1993), we found that memory recall in the model was best for the ISI-8, then ISI-64, then ISI−512 at the early RI (Figure 6C; both $p < .001$). For the RI-2048 conditions, model recall followed a ISI-512 > ISI-64 > ISI-8 (all $p < .001$), and for the Scram RI conditions,

Figure 6

Simulations Showing Spacing Effects at Extremely Long RIs



Note. (A) Data replotted from "Maintenance of Foreign Language Vocabulary and the Spacing Effect," by H. P. Bahrick, L. E. Bahrick, A. S. Bahrick, and P. E. Bahrick, 1993, *Psychological Science*, *4*(5), 316–321 (https://doi.org/10.1111/j.1467-9280.1993.tb00571.x). (B) Our modeling paradigm, wherein the same ISI was used between all training epochs, and RIs were included at a short and long delay as well as after scrambling the temporal context vector. (C) Modeling results. In (A) and (C), the optimal ISI for the same RI is marked with an asterisk. Data from simulations were plotted as mean ± SEM across runs in the model (in some cases, error bars are too small to visualize). RI = retention interval; ISI = interstimulus interval; Rep = repetition; SEM = standard error measurement. See the online article for the color version of this figure.

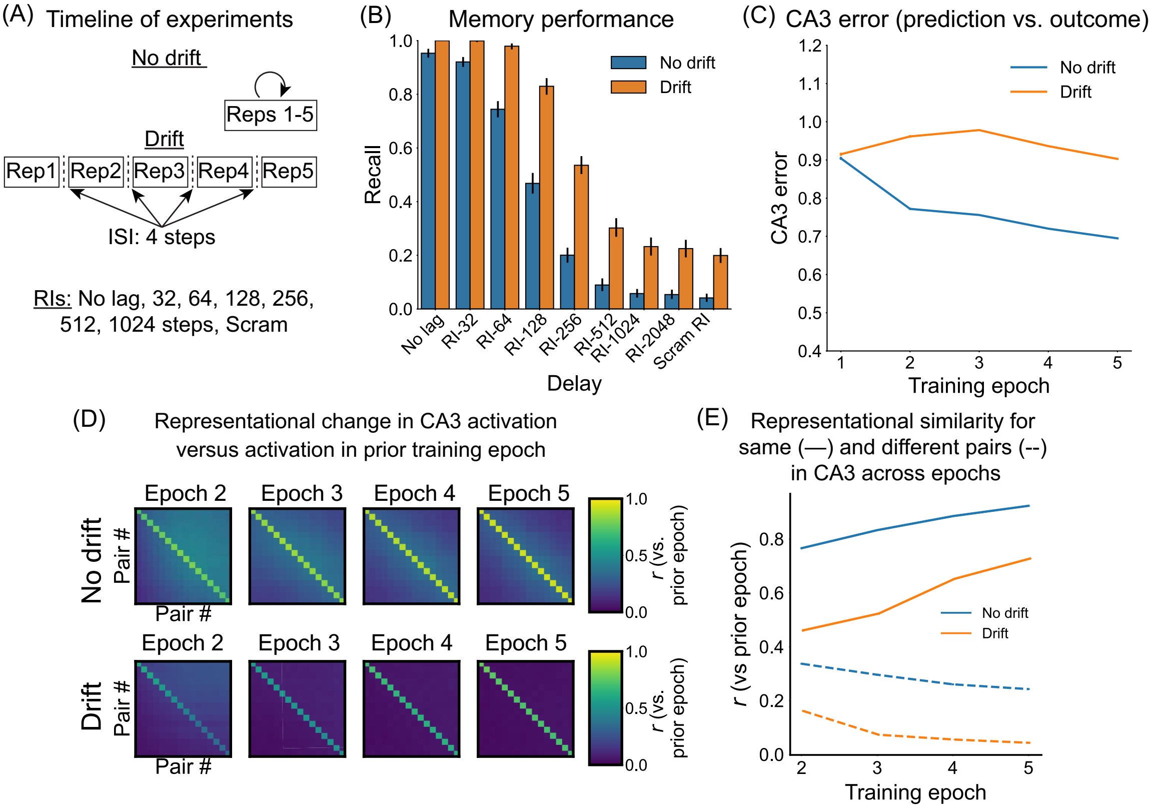
model recall followed a ISI-512 = ISI-64 ($p = .95$) > ISI-8 pattern (other contrasts, $p < .001$). These results—especially those with a scrambled temporal context—suggest that spacing benefits in HipSTeR go beyond strengthening connections within the temporal context layer, but also benefit *decontextualized* connections between the cues and targets themselves. As we will discuss below in investigating the mechanisms of learning in the model, these benefits accrue during later training epochs based on greater errors between stored and current temporal context vectors.

Drift Versus No Drift: Learning Mechanisms

Having established that the model can reproduce a number of spacing-related effects on memory, we next probed the mechanisms underlying these effects in HipSTeR. To simplify our initial investigations, we first contrasted performance in the model with no drift between training epochs—the canonical way in which neural networks are trained—against performance with modest drift, specifically four time steps + the drift during the list itself (ISI-4). We tested the model with various RIs after the final training epoch, including a No Lag RI and a condition with a scrambled temporal context (Scram RI; Figure 7A). As expected, a two-way ISI × RI ANOVA revealed a significant main effect of RI, $F(8, 1782) = 2135.5$, $p < .001$, in line with forgetting (Figure 7B). Importantly, the Drift condition outperformed the No drift condition in recall generally, $F(1, 1782) = 1356.6$, $p < .001$, and did so increasingly at longer RIs, as shown by an interaction, $F(8, 1782) = 42.2$, $p < .001$. These results suggest that drift during training makes the learned

Figure 7

Effects of Imposing Neural Drift Between Model Training Epochs



Note. (A) The timeline of experiments includes one regimen with five training epochs composed of identical temporal contexts for each paired associate trial (No Drift) and another with modest drift between each epoch (four time points). Each regimen occurred in eight training conditions, followed by retention intervals consisting of either no lag (e.g., the same temporal context as the final training trial) or six lags (32 to 1,024 time points). (B) Model memory recall was plotted for No Drift and Drift conditions across all retention intervals. Drift improved memory performance, especially after long retention intervals. (C) Error between the first quarter of the theta cycle in CA3 (the model's prediction based on the stored pattern) versus the final quarter (the "plus" phase) remained higher across training for the Drift than the No drift condition. (D) Representational stability across training epochs, measured by correlating CA3 unit activation patterns for each trial from a given training epoch against each trial of the prior epoch, was plotted for both the No Drift (top) and Drift (bottom) conditions. Trials along the diagonal represent the same pairs across epochs, whereas off-diagonal trials represent different pairs. (E) Across training, representations from epoch to epoch increased for the same pairs (solid lines) and decreased across pairs (dotted lines). The No Drift regime produced higher same- and across-pair representations than the Drift condition. Data in B, C, and E are shown as mean ± SEM across runs in the model (in some cases, error bars are too small to visualize). ISI = interstimulus interval; RI = retention interval; Rep = repetition; CA3 = cornu ammonis 3; SEM = standard error measurement. See the online article for the color version of this figure.

representations more resistant to further drift. Curiously and unexpectedly, the Drift condition even slightly bested the No drift condition at the No lag RI, which featured the exact same temporal context vectors used for all five training epochs in the No drift condition, $t(198) = 7.1$, $p < .001$. Although these results were unexpected, they are reminiscent of the impoverished learning that occurs in human behavior when two (massed) learning trials are presented with no delay (Benjamin & Tullis, 2010; Thios & D'Agostino, 1976; Xue et al., 2011).

We next sought to characterize mechanistic differences between HipSTeR trained in the Drift and No drift conditions. First, to gain insight as to how EDL differs between these conditions, we measured error in area CA3, which acts as a convergence zone between the direct pathway from ECin and the indirect pathway from DG. Recall that training trials in HipSTeR have four quarter stages resembling distinct theta phases. Error in area CA3 occurs because information arrives at different phases from different sources: During the first quarter (Q1), information arrives from ECin via the ECin → CA3 pathway, whereas in later quarters, information arrives from ECin → DG → CA3. As a result, the signal from DG effectively teaches and adjusts direct weights from EC → CA3 accordingly. Therefore, we were especially interested in the error contrasting CA3 Q1 and Q4 (plus phase) activity, which drives EDL in this area. In line with typical learning effects when using repetitions of identical input patterns, this error signal quickly abated across training epochs in the No drift condition (Figure 7C). Conversely, because temporal context patterns continually changed across training epochs in the Drift condition, this error remained high, driving greater subsequent weight changes. These patterns were supported by results from a mixed, two-way ANOVA on CA3 error, with condition (Drift vs. No Drift) and learning epoch (1–5) as factors. This ANOVA revealed a main effect of condition, $F(1, 9990) = 5042.5$, $p < .001$, a main effect of epoch, $F(4, 9990) = 952.1$, $p < .001$, and a significant interaction, $F(4, 9990) = 623.6$, $p < .001$.

Next, we asked a related question of how drift affects learning-related representational change on successive training epochs to ask how well learning drives pattern separation across pairs. To do this, we measured activity patterns across CA3 neurons at the end of the trial (Q4), and we correlated each pattern against all pairs in a given training epoch against the prior epoch (e.g., Epoch 2 vs. Epoch 1). To test this, we ran a three-way condition (Drift vs. No Drift) × Learning Epoch (2–5) × Pattern Status (same vs. different) ANOVA. As expected, activity patterns were more similar for the same than different pairs in both the No drift and Drift conditions, as shown by the bright diagonal line in Figure 7D and revealed by a main effect of pattern status, $F(1, 1584) = 153796.0$, $p < .001$. However, the nature of these correlations differed markedly between the conditions, as revealed by a main effect of condition, $F(1, 1584) = 28344.3$, $p < .001$. The main effect of the learning epoch, as well as every two-way interaction and the three-way interaction, were all significant (all $F > 114$, $p < .001$). In the No drift condition, similarity with the same pair was very high across training epochs, and similarity was substantially lower in the Drift condition (Figure 7E). This result likely reflected the fact that the input pattern was identical across epochs without drift, whereas it always slightly differed with drift. Intriguingly, however, representational similarity for a given pair against different pairs in the list also markedly differed between the conditions. In this case, similarity to different cues was much lower for the Drift condition throughout training, suggesting there was more separation between patterns in this condition.

Greater Drift Between Training Epochs Drives Temporal Abstraction and Decontextualization

In the preceding section, we demonstrated that drift (relative to no drift) benefited long-term memory, produced higher training-related errors, and drove pattern separation between memories. Intriguingly, we also found that drift benefited memory when temporal context vectors were scrambled at the test, suggesting it improved direct connections between the cues and targets. These results beg the question: does greater drift between learning events benefit memory by strengthening longer and longer time scale representations (temporal abstraction; Toppino & Gerbier, 2014), does it benefit memory by improving cue–target connections (decontextualization), or both (Figure 1)?

To address this question, we ran simulations with a large range of ISIs and RIs. Specifically, we included 11 ISIs between training epochs [nine Drift conditions with 2–512 time steps of drift, plus a No drift condition with no drift between learning epochs and a Scrambled ISI condition with a random temporal context vector between each learning epoch] and nine RIs [seven RIs after the last training epoch, from 32 (2^5) to 2048 (2^{11}) time steps, plus a No Lag condition using the final training temporal context and a Scrambled condition with a random temporal context at test]. We will present the results by speaking generally about relatively short and long ISIs and RIs, and we will summarize and interpret this entire subsection below.

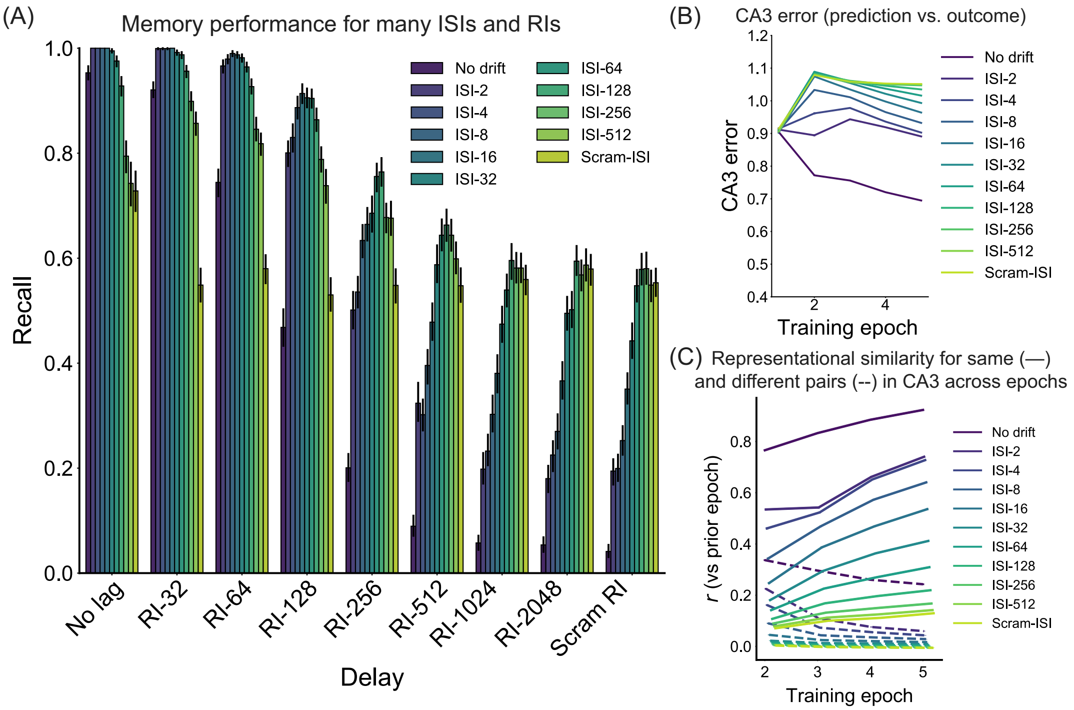
In line with our prior results and canonical spacing effects (Benjamin & Tullis, 2010; Cepeda et al., 2006), the relationship between ISI and RI was nonmonotonic, and memory was best for small ISI conditions at short RIs (and the No RI Lag condition) and large ISI conditions at long RIs (and the Scrambled RI condition; Figure 8A). This was supported by a two-way ANOVA on memory performance, with ISI × RI as factors, which showed significant main effects for both factors and a significant interaction (all $F > 96$, $p < .001$). Moreover, even in the Scrambled RI condition, longer ISIs produced better recall, suggesting more drift produced more decontextualized, direct cue–target benefits. However, the forgetting effect occurred gradually across RIs, such that the middle ISI condition had better memory recall than the shortest- and longest-ISI conditions by RI-128 (both, $p < .001$). Importantly, in the later RIs, such as RI-512, the recall was best in the moderate-ISI conditions and still better than in RI-2048 and Scrambled RI conditions (both, $p < .001$). This suggests that, in addition to direct cue–target strengthening, there was long-term strengthening of slow-drifting temporal context units, though the benefits of this became eliminated by further temporal context drift (at longer RIs). This constitutes evidence for long-term temporal abstraction as a process apart from complete decontextualization.

In accordance with findings showing greater CA3 error in the Drift than No drift condition in the preceding section, we found a main effect of condition, $F(10, 5445) = 1128.6$, $p < .001$, such that CA3 error increased with ISI (Figure 8B). We also found a main effect of epoch, $F(4, 5445) = 162.6$, $p < .001$, and a significant interaction, $F(40, 5445) = 160.1$, $p < .001$.

Additionally, as in the preceding section, we ran a three-way, ISI × Learning Epoch (2–5) × Pattern Status (same vs. different) ANOVA. We again found that all main effects, two-way and three-way interactions, were significant (all $F > 272$, all $p < .001$).

Figure 8

Effects of Different Amounts of Spacing Between Training Epochs on Model Performance



Note. (A) Different amounts of spacing influenced memory performance depending on the retention interval: shorter spacing benefited memory best at short retention intervals, whereas greater spacing benefited memory best at long retention intervals. (B) Across training, CA3 errors remained higher across epochs with greater spacing. (C) Across training, representations from epoch to epoch increased for the same cues (solid lines) and decreased across cues (dotted lines). Greater spacing produced lower same- and across-pair representational stability values. ISI = interstimulus interval; RI = retention interval; CA3 = cornu ammonis 3. See the online article for the color version of this figure.

Notably, greater ISIs produced lower within-pattern similarity but also lower across-pattern similarity throughout training (Figure 8C).

Critically, to address the question of temporal abstraction, we measured the mean temporal context pool weights between ECin → CA3 for each ISI against the No drift condition. A two-way ISI by temporal context pool interaction on mean weights produced the main effects of both factors and a significant interaction (all $F > 444$, all $p < .001$). For the interaction, we found a full crossover effect between the temporal context pool and ISI: the fast-drifting pools had the greatest mean weights in the short-ISI conditions, whereas the slow-drifting pools had the greatest mean weights in the long-ISI conditions (Figure 9).

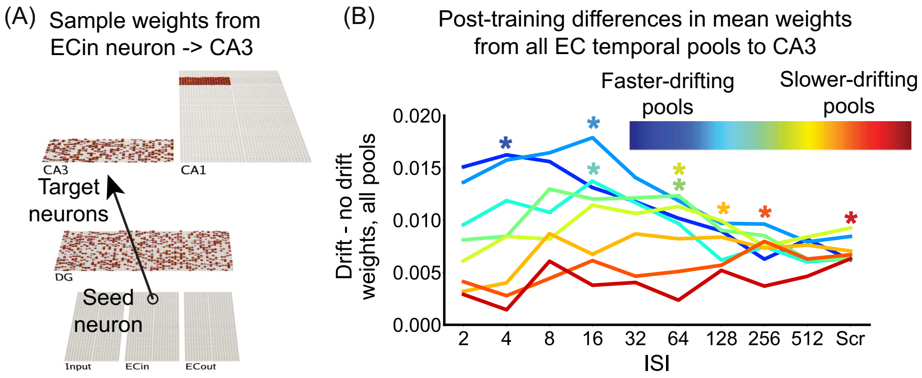
We now offer a cumulative explanation of these results. The short-RI advantage for short-ISI conditions occurs because the fast-drifting temporal context vectors can still offer cue support to the memory after short RIs, so strengthening these vectors benefits recall at these time points. For short-ISI conditions tested after long RIs, strengthening the fast-drifting temporal context pool weights does very little to support recalling the memory because the test patterns have drifted far away from the final training pattern; this condition has caused the model to effectively overfit to a local temporal context, and therefore forgetting occurs more quickly. The long-ISI conditions are inferior at short RIs because so much drift has occurred between training epochs that, while they can strengthen the slow-drifting temporal context pool weights, they do not strengthen many of the fast-drifting temporal context weights as well; these fast-drifting weights are effectively scrambled by the time new training epochs occur, so each new training instance effectively strengthens a new random subset of weights. Nevertheless, the slow-drifting vectors retain some overlap across training epochs, and their corresponding pool weights become strengthened. Therefore, the slow-drifting pools support memory recall better in long-ISI conditions at long RIs than in short-ISI conditions. This temporal abstraction process confers advantages for the long-ISI conditions at medium-to-long RIs before the slow-drifting pools have also drifted to a chance level of overlap. Finally, greater error also results in more decontextualized cue–target associations because the same ECin units will be activated for each cue and target, and their corresponding weight changes in HC will be stronger due to the higher error in the network layer. This produces spacing benefits at the longest RIs and when the temporal context pools have become fully scrambled (Figure 1).

Eliminating Learning in Specific Pathways Reveals Dissociable Learning Mechanisms Within the HC

We have thus far outlined two primary learning mechanisms that can support spaced learning: temporal abstraction and

Figure 9

Temporal Abstraction in HipSTeR



Note. (A) Weight strengths from a sample seed neuron from ECin → CA3 units are shown. (B) Posttraining differences in mean weights between each ISI against the No Drift condition are shown for each temporal context pool. Asterisks indicate the ISI with the strongest mean weight for each pool. Across spacing conditions, greater spacing produced weaker weights for the shorter timescale pools and stronger weights for the longer timescale pools. HipSTeR = Hip-pocampus with Spectral Te-mporal Representations; ECin = entorhinal cortex input layer; ISI = inter-stimulus interval; DG = dentate gyrus; ECout = entorhinal cortex output layer; EC = entorhinal cortex; CA3 = cornu ammonis 3; CA1 = cornu ammonis 1. See the online article for the color version of this figure.

decontextualization. To examine how different pathways and learning mechanisms in HipSTeR produced these results, we next performed analyses involving turning off learning or only EDL in specific pathways. We contrasted the full HipSTeR model in the preceding section against models that were identical except for the following changes: ECin → DG (no learning), ECin → CA3 (no learning), ECin → CA3 (no EDL, but Hebbian learning was present), ECin → CA1 (no learning), CA3 → CA3 (no learning), and CA3 → CA1 (no learning). We will present these results by their increasing relevance for elucidating learning mechanisms.

As expected, turning off learning between CA3 → CA1 abolished learning completely in the model, as it is the only connection from the trisynaptic and disynaptic pathways (Figure 10, top center). Conversely, turning off learning from ECin → CA1 had almost no effect on performance (Figure 10, top right); however, at later intervals, this model actually outperformed the full HipSTeR model in short-ISI conditions tested at long RIs (more on this below). This was supported by a two-way ISI × RI ANOVA on memory differences between the control and lesion models, which showed the main effects of both factors and, critically, a significant interaction (all $F > 5$, all $p < .001$). Turning off learning between the recurrent CA3 → CA3 pathway produced moderate impairments relative to the full model that increased with drift (Figure 10, top left), which was supported by a similar ANOVA showing a main effect of RI, $F(4, 780) = 26.7$, $p < .001$, no main effect of ISI, $F(3, 780) = 0.4$, $p = .54$, and a significant interaction, $F(12, 780) = 25.3$, all $p < .001$.

We saw the strongest and most interesting dissociation between models without ECin → CA3 learning and without ECin → DG learning. Turning off learning completely from ECin → CA3 strongly affected cue–target learning, such that the memory became hypersusceptible to drift (i.e., performance decreased more quickly with drift) relative to the full HipSTeR model, Figure 10, left middle; main effect of RI: $F(4, 780) = 230.4$, $p < .001$; no main effect of ISI: $F(3, 780) = 1.4$, $p = .54$; interaction: $F(12, 780) = 533.4$, all $p < .001$. This was similarly the case when we turned off EDL but kept Hebbian learning in this pathway (Figure 10, bottom left; both main effects and interaction: $F > 34$, $p < .001$); the differences between the complete nonlearning and the no EDL ECin → CA3 models were significant, but they were quantitatively small (Figure 10, bottom center; both main effects and interaction: $F > 142$, $p < .001$). These results suggest that EDL from ECin → CA3 strongly drives the drift-resistant, temporally abstracted, or decontextualized part of the memory, directly linking cue and target. Critically, they also reveal a novel mechanism by which decontextualization—generally thought to be a process confined to the neocortex (Hasselmo, 2005; Winocur et al., 2010)—could occur within the HC itself.

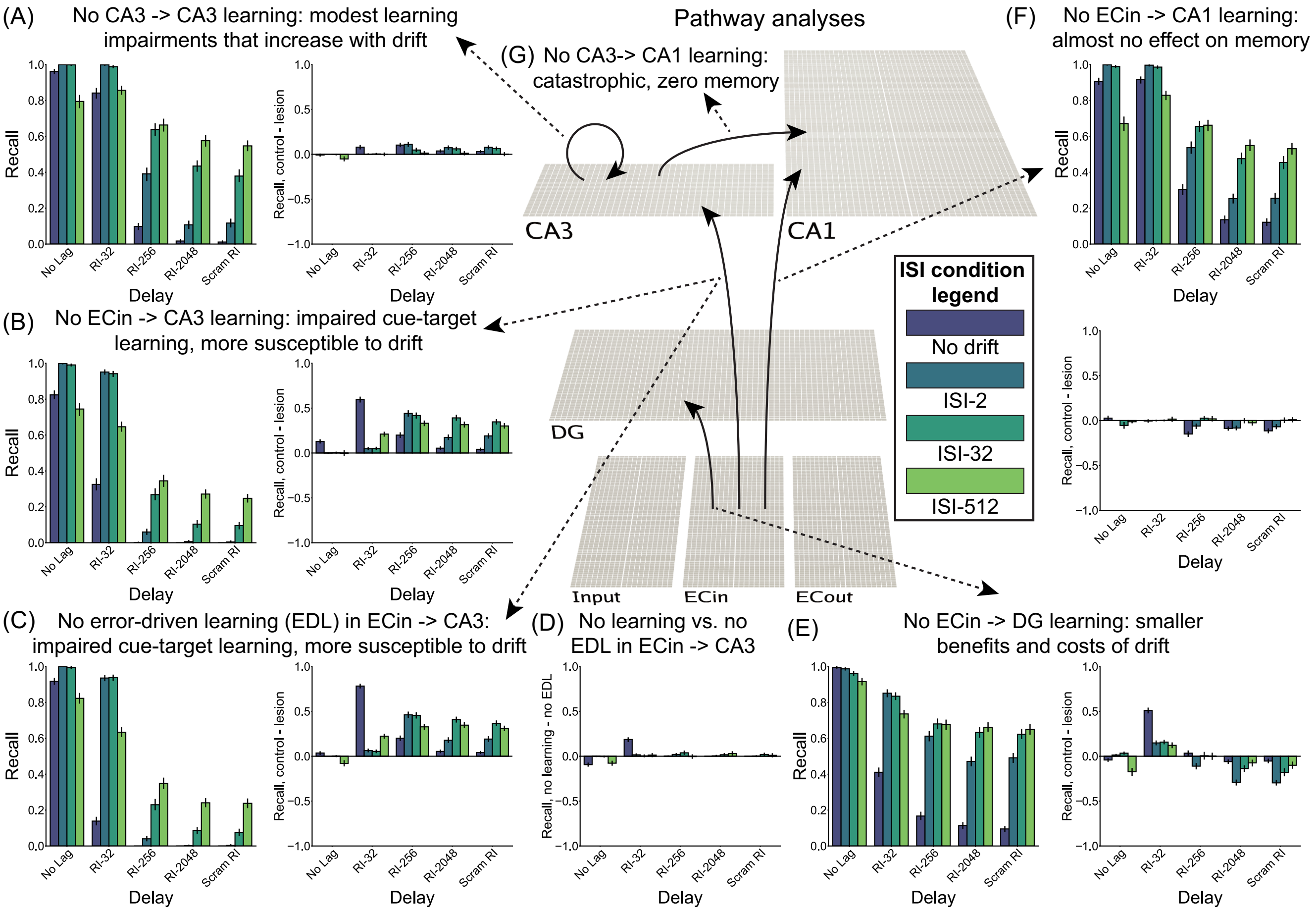
As opposed to the ECin → CA3 pathway, turning off ECin → DG learning rendered the memory largely resistant to drift. Relative to the full HipSTeR model, this model mostly lacked the benefits conferred by the intact temporal context at short RIs, which have temporal contexts that have not yet drifted away from their training contexts (Figure 10, bottom right; both main effects and interaction: $F > 26$, $p < .001$). On the other hand, without learning in this pathway, error in CA3 remained higher during training, which helped to strengthen the direct cue–target aspect of the memory (dependent on ECin → CA3). As a result, this model paradoxically outperformed the full HipSTeR model at long RIs (and the very weak benefits at long RIs in the nonlearning ECin → CA1 model likely occurred for similar reasons). In sum, the more direct ECin → CA3 pathway produced the spacing benefits of temporal abstraction and decontextualization that make memories more drift-resistant, whereas the ECin → DG pathway produced transitory benefits that continued to rely on temporal context, the scale of which depended on the amount of drift during training.

Comparing Multiscale to Uniform Drift

Many prior temporal context models posit that drift occurs across multiple time scales (e.g., Liu et al., 2019), which is supported by

Figure 10

Removing Learning in Different CLS Pathways Revealed Dissociable Learning Mechanisms

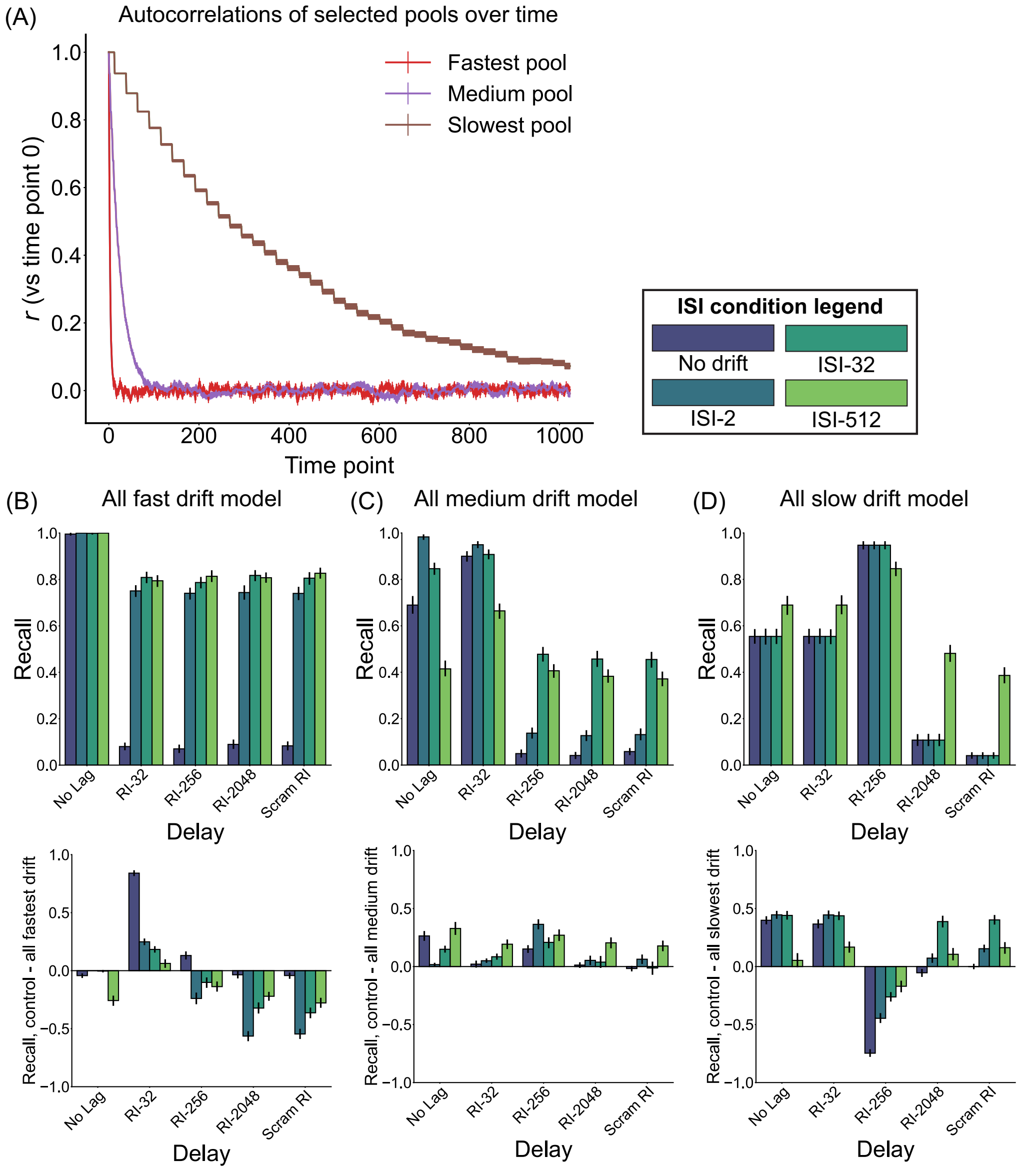


Note. We plotted HipSTeR model architecture in the center, with affected pathways shown as arrows connecting results on the periphery. For simplicity, we plotted retention interval conditions for only four spacing conditions: No Drift, ISI-2, ISI-32, and ISI-512, representing no, short, medium, and long spacing. (A) Preventing CA3 → CA3 learning in the model left performance generally preserved (left). Comparing this with the full HipSTeR model shows only modest improvements for the full model that increase with greater drift (right). (B) Preventing learning from ECin → CA3 impaired cue–target learning, making performance highly susceptible to drift (left). Comparisons with the full model showed nearly uniform performance differences across retention intervals (right). (C) Specifically, preventing error-driven learning (EDL) in this pathway produced similar results, both on its own (left) and when compared with the full model (right). (D) Directly comparing the full- and Hebbian-only models revealed minimal differences. (E) Preventing learning from ECin → DG preserved cue–target learning and created a relative resistance to neural drift: Benefits of the full model are present with short retention intervals but are reversed at longer retention intervals, such that performance is actually superior in this model versus the full model. (F) Preventing learning from ECin → CA1 left memory largely intact (top), as shown by negligible differences when comparing it with the full model (bottom). (G) Preventing CA3 → CA1 learning abolished memory in all conditions and at all retention intervals (unshown). CLS = complementary learning systems; HipSTeR = Hip-pocampus with Spectral Te-mporal Representations; ISI = interstimulus interval; ECin = entorhinal cortex input layer; DG = dentate gyrus; CA3 = cornu ammonis 3; CA1 = cornu ammonis 1; RI = retention interval. See the online article for the color version of this figure.

various forms of neurobiological evidence (e.g., Tsao et al., 2018; Umbach et al., 2020). Nevertheless, it is worthwhile to assess the advantages and disadvantages of a system employing uniform as opposed to multiscale drift. In the next three simulations, we contrasted our results across numerous ISIs and RIs from our multiscale (control) model against identical models using uniform drift for all eight temporal context pools. We chose three drift rates that were equal to the fastest, slowest, and medium-speed pool from our control model, and we plotted the results alone and against our control model (Figure 11). For statistics, we ran two-way, ISI × RI ANOVAs on the memory differences for each model against the control model. First, for most ISIs and RIs in the all-fast drift model, the pattern was completely new upon each training instance because the autocorrelated pattern reaches a value of 0 after approximately 16 time steps (Figure 11A; both main effects and interaction: $F > 18$, $p < .001$). This means that the amount of drift was a much smaller factor, so most of the results after the No Lag RI were equivalent (Figure 11B). With five training epochs, the model actually outperformed the multiscale model after modest-to-long delays because the high error (from having essentially random context patterns on each training instance) drove such strong decontextualization that the memories were robust to drift. However, at short

Figure 11

Simulations Contrasting Multiscale Drift Against Uniform Drift



Note. (A) Autocorrelations for the three uniform drift rates (equivalent to Pool 1, 5, and 8 in the multiscale model, respectively). (B–D) Results from models with the fastest, medium, and slowest uniform drift rates (top) and direct contrasts between the multiscale (control) model against these models (bottom). ISI = interstimulus interval; RI = retention interval. See the online article for the color version of this figure.

delays like RI-32, the multiscale model outperformed the all-fast model due to a lack of cue support in the all-fast model.

Next, we will analyze the all-slow model (Figure 11D). In contrast to the all-fast model, after many ISIs and RIs, the temporal context pattern was almost completely intact (both main effects and interaction: $F > 20$, $p < .001$). When the ISI was low (No drift or short ISI conditions, such as ISI-32), all cue–target pairs from the five epochs were trained on almost exactly the same temporal context pattern. Note that this differs somewhat from the 'No Drift' condition in prior simulations in that the drift here is so slow that there is hardly any within-epoch drift in addition to little between-epoch drift at short ISIs. This means that the retrieval context after a short delay (such as with no lag or at RI-32) resembles the pattern of the final training epoch—and there will be massive interference among list items. Interestingly, the model actually benefited from some drift to "unstick" the memory from having basically the same pattern across all epochs, which, when it remained the same for all pairs within the list, caused interference among the pairs. These performance benefits arose with either moderate RIs (e.g., RI-256), which were long enough to prevent some of the unsticking that occurred at low RIs but not so long that the very slow-drifting patterns nonetheless drifted away (e.g., RI-2048). The benefits can also occur with more drift during learning epochs (e.g., ISI-512). In these cases, the model learned different temporal patterns in each instance and did not repeatedly strengthen the same (ultimately interfering) weights. The model was inferior to the multiscale model at most intervals because of either this interference (at short RIs) or reduced EDL because the pattern remained intact for longer (at long RIs). However, it was superior to the multiscale model at a moderate RI like RI-256, indicating that there is some RI for which it is optimized. Ultimately, the slow drift model allowed patterns to be maintained for long periods of time but suffered from interference due to poor within-list drift. Finally, the all-medium model fell between these two extremes, performing closest to the control model of the three simulations but showing deficits with very short or long lags (Figure 11C; both main effects and interaction: $F > 17$, $p < .001$).

It is never certain during learning when information will be relearned or need to be retrieved. Therefore, we interpret these results to indicate that multiscale drift does not optimize for any particular RI but rather a range of potential RIs, balancing learning and long-term memory maintenance while allowing for relearning benefits according to the temporal frequency of the information.

Decontextualization in Other Paradigms

To the best of our knowledge, conceptualizing spacing effects as temporal decontextualization is novel. Therefore, our final simulations aimed to bridge our modeling framework to other studies falling under the umbrella of decontextualization, which has used paradigms with environmental, task, or background (pictorial) contexts (Butler et al., 2017; Glass, 2009; Maskarinec & Thompson, 1976; S. M. Smith et al., 1978; S. M. Smith & Handy, 2014, 2016; S. M. Smith & Rothkopf, 1984; Soderstrom & Bjork, 2015; Trask & Bouton, 2018; Zawadzka et al., 2021). These final simulations thus demonstrate that our general modeling approach can also capture other decontextualization effects.

In decontextualization paradigms, the contexts are either constant or variable before final tests take place in novel contexts. In S. M. Smith et al. (1978), subjects learned words in Room A and either practiced recalling the words in Room A or in Room B in a second session before taking a final test in Room C. Critically, memory performance in the variable room (relative to the constant room) was worse during practice but better at the final test in Room C, suggesting the environmental variability decontextualized the memory. Imundo et al. (2021) used a similar method but had subjects either restudy or retrieve lists (without feedback) during the second session. They found variability benefits, but only during restudy, whereas retrieval (without feedback) likely relied on successful retrieval for benefits. In S. M. Smith and Handy (2016), subjects learned word pairs against background contexts that were unique for each pair and either remained constant across five practice trials or varied across trials. A final test was given for all pairs with no background context. As in S. M. Smith et al. (1978) and Imundo et al. (2021), cued recall for pairs in the variable (relative to the constant) condition was worse during learning but better at the final test, suggesting trial-specific pictorial context variability also decontextualized memories. Following the logic outlined in the preceding sections, these findings could arise because variable contexts offered less cue support to memories during learning, and this, therefore, led to the more direct strengthening of the cue–target elements that were in common across training epochs. These changes would then support the memory when tested in novel contexts. Note that, for the purposes of our simulations, S. M. Smith et al. (1978) manipulated context at the epoch level, whereas S. M. Smith and Handy (2016) manipulated context at the trial pair level. Note also that if the test in these paradigms occurred instead with the original training context, pairs in the constant (relative to the variable) context should produce better memory, in line with some of both their results (S. M. Smith et al., 1978) and findings akin to context-dependent memory (Godden & Baddeley, 1975). These effects should arise because weights related to the original training context should be repeatedly strengthened across training (Cox et al., 2021; Estes, 1955; S. M. Smith & Handy, 2014, 2016).

To simulate these results, we kept the same HipSTeR model architecture and slightly modified the inputs (Figure 12A). The inputs included the same number of cue and target pools and the six fastest drifting temporal context pools (out of eight). Here, instead of the two slowest drifting temporal context pools, we added two other context pools. In our epoch-wise simulations (representing environmental context, as in S. M. Smith et al., 1978), the context vectors were identical throughout a learning epoch. Across epochs, they either remained the same (constant) or changed for each epoch (variable). At the final test, these pools had either the context vectors from the first learning epoch (old) or a new vector (vew). In our trial-wise simulations (representing background pictorial contexts, as in S. M. Smith & Handy, 2016), the context vectors were unique for each pair. Across epochs, they either remained the same for that specific pair (constant) or changed on each epoch (variable). At the final test, these pools had either the same specific context vectors for that pair from the first learning epoch (old) or a new vector (new). In all experiments, we used a temporal ISI of two steps and an RI of 512 steps.

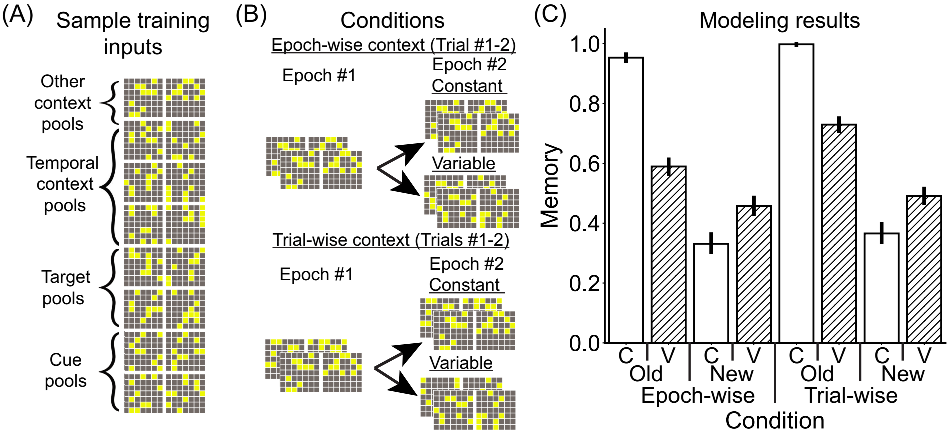
We analyzed these data with a three-way Training Type (epoch-wise vs. trial-wise) × Test Context (old vs. new) × Training Variability (constant vs. variable) ANOVA on model memory. We found that all three main effects, all three two-way interactions, and the three-way interaction were all significant (all $F > 40$, $p < .001$). The main effect of training type showed that overall memory performance was higher in the trial-wise than epoch-wise simulations. This likely occurred because the more specific context cues resulted in more efficient learning with less incidental interference resulting from the same context applied to all pairs of a list (Nairne, 2002; Figure 12B). The other main effects indicated that memory was better overall with constant training and old contexts presented in the text, but note that the interactions are of more interest here. In line with prior results and our predictions, in both epoch-wise and trial-wise simulations, final test performance was superior in the constant relative to the variable condition when tested with the old context ($p < .001$), showing unsurprisingly that cue support from training aids memory (Cox et al., 2021). More critically, however, performance was superior in the variable relative to the constant condition when tested with a new context ($p < .001$), suggesting that contextual variability led to decontextualized memory traces that allowed for successful memory recall even with new contexts.

Discussion

By implementing drift and EDL within HipSTeR, a biologically plausible model of the EC–HC network, we simulated forgetting

Figure 12

Simulations of Decontextualization in Other Paradigms



Note. (A) For these simulations, we modified HipSTeR model inputs slightly by replacing two temporal context pools with other context pools. (B) Other context pools were either constant within an epoch (top) or unique for each cue–target pair (bottom). (top) Epoch-wise contextual changes were always constant within an epoch (top left) and either constant across all epochs or variable across epochs (top right). (bottom) Trial-wise contextual changes were always different for every trial, even within an epoch (bottom left), and were either in a constant order in each epoch or variable for each trial (bottom right). (C) Modeling results showed that, while performance was superior in the constant training context when the constant (old) context was given at the test, performance was superior in the variable training context when a new context was given at the test. HipSTeR = Hip-pocampus with Spectral Te-mporal Representations; C = constant training context; V = variable training context. See the online article for the color version of this figure.

and a wide array of spacing effects from the cognitive psychology literature. First, we ran simulations that replicated specific spacing effect principles. We found that the optimal spacing ISI depended on the RI before the final test (Cepeda et al., 2008), that relearning after long temporal delays overrode smaller differences due to initial spacing (Rawson et al., 2018), that different spacing schedules produced modest differences (Küpper-Tetzel et al., 2014) but that absolute spacing was the most critical factor (Karpicke & Bauernschmidt, 2011), and that spacing can produce direct benefits between cues and targets at extremely long RIs—intervals when there is little (to arguably no) resemblance between the temporal contexts during training and at test (as in the 5-year follow-up test in Bahrick et al., 1993).

Next, we probed the mechanisms producing these effects in HipSTeR. The very presence of temporal context drift (relative to no drift) resulted in better recall performance, greater error in CA3 in later training epochs, and greater representational dissimilarity between different CA3 activation patterns. Further analyses investigating a full spectrum of drift values between training epochs showed that greater spacing produced more temporal abstraction (as shown via stronger mean weights in slower temporal context pools from ECin → CA3) and decontextualization (as shown via better recall when temporal context vectors were scrambled; Figure 1). These analyses also showed that massed learning can be superior in the short term because there is preferential strengthening in the faster drifting temporal context vectors, which can still provide cue support for the memory after low drift.

Turning off learning in various pathways in HipSTeR showed a stark dissociation between a pathway that continued to support contextualized memories and benefited the model under conditions of low drift between the training and testing context (ECin → DG → CA3) and one that supports decontextualization (ECin → CA3). The latter mechanism is especially novel because decontextualization has generally been attributed to the neocortex (Winocur et al., 2010; Yassa & Reagh, 2013), but here we show how it can arise in the HC. Comparing our model using multiscale drift against a set of uniformly drifting alternatives showed that, while the uniform models performed better when testing after some particular RIs, the multiscale model seemed to balance memory performance for a range of potential RIs. Finally, because conceptualizing spacing effects as arising partly due to decontextualization is novel, we linked our results with more canonical decontextualization effects in the literature. Using an identical model architecture and similar inputs, HipSTeR captured classic decontextualization effects, showing that variable encoding contexts lead to better memory when tested in novel contexts.

Comparisons With Other Spacing Effect and Encoding Variability Theories and Models

Many articles argue that spacing effects could arise from some combination of two processes: encoding variability, which determines the amount of strengthening, and study-phase retrieval, which determines the likelihood of strengthening (Benjamin & Tullis, 2010; Küpper-Tetzel & Erdfelder, 2012; Landauer, 1969; Mozer et al., 2009; Raaijmakers, 2003; Smolen et al., 2016; Walsh et al., 2018). Encoding variability theory asserts that re-encoding benefits increase with more variable encoding contexts (such as after greater temporal lags) because there will be a greater diversity of contextual elements added to the memory, resulting in more routes to retrieval

(Estes, 1955; Maddox, 2016; Mozer et al., 2009; S. M. Smith et al., 1978; S. M. Smith & Handy, 2014). Study-phase retrieval asserts that in order to benefit from spacing, subjects must retrieve the episode from encoding (Thios & D'Agostino, 1976). Ultimately, because encoding variability would produce benefits that increase monotonically with spacing and because study-phase retrieval would be best immediately after encoding and, therefore, would produce antispacing benefits as a solo mechanism, the two have been proposed in combination to capture the nonmonotonic nature of spacing effects (Benjamin & Tullis, 2010). From these foundations, a question arises as to which representations become strengthened and why. One intriguing theory is that there is a temporal abstraction process that depends on the ISI (Mozer et al., 2009; Toppino & Gerbier, 2014). That is, longer ISIs may preferentially strengthen aspects of the memory trace that have more long-term stability. In particular, Mozer et al. (2009) created a multitemporal scale neural network model in which strengthening preferentially occurred in longer-term representations as a function of the error between encoding contexts along shorter time scales, such that a failure to support the memory along shorter time scales forced strengthening along a longer time scale.

How do these theories and models square with results from HipSTeR? We found that encoding variability (in the form of temporal context vector differences) benefited memory at the longest RIs via two EDL-driven mechanisms (Figure 1). Like Mozer et al. (2009), greater ISIs caused temporal abstraction or greater relative strengthening in slower drifting temporal pool weights. Greater ISIs also allowed for better intact memory in the face of the complete decontextualization of fully scrambled context vectors. There is, therefore, a critical difference between how variability produces strengthening in encoding variability theory and in HipSTeR: encoding variability theory suggests that strengthening occurs because the same memory accrues more contextual routes to retrieval, whereas our model suggests it occurs because greater error produces strengthening of the relevant temporal context pool weights and/or cue–target weights. That is, our results suggest that gaining more unspecified contextual routes to the memory may not be ultimately that helpful since it is unclear, given that drift will occur randomly, how or why any of them should be meaningfully activated during retrieval after further drift has occurred.

Instead, drift in HipSTeR strengthened the weights of slower drifting units that were more likely to be active at retrieval or decontextualized cue–target weights that were fully drift-resistant. We believe this latter explanation seems almost necessary to explain results at extremely long RIs, such as the 5-year RI from Bahrick et al. (1993) or results in other paradigms showing retention across decades (Maxcey et al., 2022). Moreover, this account can be considered even more straightforwardly in our final environmental/pictorial decontextualization paradigm simulations. We tested these models with novel context vectors, and there was no reason why these (random) vectors should have activated any of the contextual elements accrued in the variable condition any more than they would have activated unhelpful elements that would have counteracted memory retrieval. Altogether, rather than encoding variability driving memory strengthening while study-phase retrieval limits the likelihood of strengthening, in HipSTeR, EDL drove strengthening via temporal abstraction and decontextualization while contextual drift produced forgetting with elapsing time.

A recent article showed that as rodents accumulated experience within the same environment across multiple days, some place cells—or cells that fire at particular locations within a spatial context—transiently entered and left the memory trace, whereas other ones sustained a stable firing location across days (Vaidya et al., 2023). Intriguingly, the proportion of sustained neurons grew over time and predicted behavioral performance in accordance with the memory trace reactivating and solidifying itself with experience. We believe that an experiment manipulating spacing and using a similar analytical approach that also measures their endurance (how long they stay in the memory trace) could help adjudicate between the predictions of encoding variability and our model. That is, encoding variability theory predicts that spacing would increase the *proportion* of new neurons in a trace upon relearning (even if they were ultimately transient and not part of the trace in the next learning session). Conversely, our model predicts that spacing would increase the number and endurance of sustained neurons in the trace. A further prediction of our model, given the principle of temporal abstraction, is that the amount of spacing should directly scale with neuron endurance.

These findings resonate with an influential theory that separates memory constructs into retrieval and storage strength (Bjork & Bjork, 1992; RBjork, 2011). Retrieval strength refers to the in-the-moment accessibility of memory and explains memory performance at a particular time, whereas storage strength refers to a latent factor referring to how well the memory is learned and explains its persistence at later moments. Note that prior researchers similarly distinguished between constructs related to current and later performance (Estes, 1955; Hull, 1943; Skinner, 1938). Mapping these constructs onto our modeling findings, we believe retrieval strength relies on momentary pattern matches that depend partially on the overlap between the learned temporal context and the temporal context at retrieval. Conversely, both temporal abstracted weights (which help the memory for longer periods of time but not necessarily forever) and decontextualized weights (which directly strengthen cue–target associations) align with the concept of storage strength.

The Organization of Time and Model Plausibility

Researchers have long been interested in the role of time in episodic memory. "Mental time travel" was a core aspect of Tulving's (1972) conceptualization of episodic memory, suggesting the reinstatement of a particular temporal frame. Countless researchers have investigated the organization of time in memory, with a core question being whether or not it is organized by chronological "time stamps" (Bradburn et al., 1987; Burt, 1992; Friedman, 1993; Hintzman, 2016; Jeunehomme et al., 2018). Indeed, contiguity effects in memory, whereby information presented nearby in time is also temporally clustered together during recall, suggest that individuals reinstate temporal contexts that guide further retrieval (e.g., Howard & Kahana, 2002; Lohnas et al., 2015). Intriguingly, temporal clustering occurs across numerous time scales, across lists, and up to days and months (Healey et al., 2019; Howard et al., 2008; Moreton & Ward, 2010; Uitvlugt & Healey, 2019; Unsworth, 2008), suggesting a scale-invariant property in support of log-spaced temporal representations (Brown et al., 2007; Howard, 2018). Moreover, reinstatement has also shown neural contiguity effects (Folkerts et al., 2018; Manning et al., 2011), and it has been proposed that another role of neural drift is to create such time stamps (A. Rubin et al., 2015).

However, there have been notable criticisms of these ideas (e.g., Friedman, 1993; Hintzman, 2016). Many free recall tasks rely on intentional encoding, whereby subjects can develop rehearsal strategies that rely on successive rehearsals, and incidental encoding drastically reduces (though does not eliminate) temporal contiguity effects (Dester et al., 2021; Healey, 2018; Mundorf et al., 2021). Moreover, temporal contiguity effects can be reduced or altered by other types of (e.g., semantic, narrative) structure (e.g., Antony et al., 2021; Bousfield, 1953; Heusser et al., 2021; Polyn et al., 2009), suggesting these effects are highly manipulable in the presence of more salient organizing characteristics. Additionally, temporal context theories might predict that successively presented cued recall trials would differ as a function of their lag at encoding, but multiple experiments have shown no effect of this lag (Osth & Fox, 2019). At the very least, these findings question the idea that we automatically encode and retrieve information as if reading from a timeline.

How can we resolve these ideas and determine to what extent drift might influence memory? We propose that time is a weak signal if not given prioritized attention. That is, the ability to reconstruct the past in a direct linear order when one is not intend to or does not encode a set of causally connected events in a linear order is poor (Dimsdale-Zucker et al., 2022) but not zero (Dester et al., 2021; Healey, 2018). Even in a rodent experiment that showed robust decoding of trial numbers from LEC neurons, when a task change eliminated the need to keep track of different trials, the ability to decode trial numbers from the neurons was reduced but not eliminated (Figure 8 experiments, Tsao et al., 2018). It is reasonable, then, to wonder what the point of such a drifting signal could be across each of these time scales. We speculate that—just like features of the environment that are represented in sensory subsystems but do not receive direct attention—the ability to use this signal relies strongly on whether one incorporates it somehow into the focus of attention (Niv et al., 2015). Though speculative, this may even be optimal from a memory standpoint, as too strong of a temporal signal may result in linking unlike events that *just so happen* to co-occur. This idea suggests that the prominence of drifting signals within a memory trace can fall under executive control, either becoming accentuated or minimized (see below the Limitation section on neural drift versus shift; DuBrow et al., 2017).

Relevance to Temporal Context Models of Memory

Dating back to Estes (1955) and Bower (1967), memory models have considered context as a pattern of activity that fluctuates over time. This family of models has shown that modeling context can capture numerous temporal memory phenomena. For instance, Mensink and Raaijmakers (1988) captured how memory interference depends on the timing of interfering information and the time of test, capturing intricate findings from the literature on proactive and retroactive interference and spontaneous recovery. Later, Howard and Kahana (2002) showed how binding learned items to a drifting context can capture the aforementioned temporal contiguity effects in free recall. These models have also generalized this idea to other types of context, such as semantic (Lohnas et al., 2015; Polyn et al., 2009) and emotional context (Talmi et al., 2019). Altogether, this family of models has simulated an impressive array of memory data.

The primary contribution of our model to these efforts is to show how memories get updated and strengthened upon repetition. Context continually drifts over time in a multiscale manner, but our model suggests that repeating information at increasing timescales will strengthen increasingly long context representations that effectively help the model generalize over time. Therefore, in HipSTeR, memory strength at a given time relies on a rich combination of drifting temporal context and the timing history of training.

Implications for the Neurobiology of Drift

The theory behind HipSTeR draws heavily from recent neurobiological evidence of drift in single neurons and neural ensembles within the medial temporal lobe. This evidence stems from two types of findings. The first involves activity from individual neurons or ensembles that ramp up or down spontaneously (Folkerts et al., 2018; Howard et al., 2012; Tsao et al., 2018; Umbach et al., 2020; Yoo et al., 2022) or in a cue- or context-evoked fashion (Bright et al., 2020; Liu et al., 2022; Tsao et al., 2018; Umbach et al., 2020; Yoo et al., 2022). This "ramping" drift, which constitutes a form of persistent activity, could occur as a result of unique, slow-adapting Ca++ currents (Egorov et al., 2002; Liu et al., 2019; Tahvildari et al., 2007; Tiganj et al., 2015). The second compares representational similarity between successive experiences of the same type after variable delays, such as neural activity patterns when animals are placed in the same environment (Bladon et al., 2019; Keinath et al., 2022; Mankin et al., 2015; Marks & Goard, 2021; Mau et al., 2018; A. Rubin et al., 2015; Rule et al., 2019; Y. Ziv et al., 2013). This drift has been attributed to fluctuations in synaptic weights (Mau et al., 2020; N. E. Ziv & Brenner, 2018) and intrinsic excitability (Delamare et al., 2023), and the difference in these drift types may be critical when considering temporal effects between those from seconds to hours against those from days to years. Therefore, although HipSTeR implements drift in a manner more resembling the former ramping type, it is worth considering in later models whether these results would hold in models with drifting synaptic weights.

One common puzzle arising within this literature on drift is how long-term memories survive amidst drift and what function drift might serve. It has been claimed (and shown via modeling) that long-term representations can continue to survive even with substantial drift (Kalle Kossio et al., 2021; Mau et al., 2020; Rule & O'Leary, 2022). Using a recurrent neural network model, Clopath et al. (2017) showed that long-term stability can be maintained with a "backbone" of stable neurons and recurrent activity. Relatedly, while there may be drift from the perspective of some external input (e.g., the environment), internal representations between neural representations and their downstream readers may remain relatively coherent and low-dimensional latent structure relatively constant, which could help to compensate for drift in earlier representations (Delamare et al., 2023; Gallego et al., 2020; Kalle Kossio et al., 2021; Mau et al., 2020; Rule et al., 2019).

Drift could also be helpful for new encoding, as it allows for greater flexibility and even forgetting of old information in an arguably rational manner. Regarding flexibility, drift allows efficient new encoding in a way that does not fully depend on or compete with old memories (Frank et al., 2018; Mau et al., 2020), as it allows for new memories to not simply be reassigned to the same, most excitable units (O'Reilly et al., 2017; Rogerson et al., 2014; Zhou et al., 2009). This would allow memories to form with sparse, efficient representations while eventually (after enough time) involving most or all of the relevant neurons in a particular region. Regarding forgetting from a rational memory standpoint, a memory encountered once or a series of experiences repeated within a short temporal frame is less likely to be relevant for the long term than those repeated more infrequently (Anderson & Milson, 1989; Anderson & Schooler, 1991; Mozer et al., 2009). That is, the likelihood of many stimuli in the environment being repeated falls off rapidly as a function of its last occurrence, a principle which has applications in such wide-ranging domains as word occurrences within *New York Times* articles and utterances from parents to children (Anderson & Schooler, 1991). Given computational constraints, it would seem rational to forget such memories to maintain both the potential to form new memories or preferentially retain those more likely to serve future needs on a timescale that resembles the frequency of encoding (see Figure 11). Therefore, we propose a novel mechanism of drift, such that it allows for useful forgetting and, when a memory becomes experienced (or reactivated) after variable delays, it allows for strengthening on a relevant timescale in the form of greater EDL-driven temporal abstraction. Additionally, to the extent that drift between training epochs constitutes a form of noise between input patterns, these results—whereby drift helps the network generalize over time—support the idea that optimal levels of noise help avoid overfitting and improve generalization in neural networks (Elman & Zipser, 1988; Hinton & van Camp, 1993; Sietsma & Dow, 1991; Srivastava et al., 2014; Tran et al., 2022).

Relationship to Theories of Hippocampal Contributions to Episodic Memory

Our results are relevant to findings relating hippocampal activity to memory of different ages. First, univariate HC activity measured using fMRI increases when recalling memories with increasing temporal lags across trials (Brozinsky et al., 2005) or days (Chen et al., 2016). Moreover, the rate of hippocampal ripples increases with the recall of older autobiographical memories (Y. Norman et al., 2021). In HipSTeR, this increase could be spurred by the amount of error encountered between the differences in temporal contexts. Second, a neural version of the encoding variability account—whereby more variability in neural representations across repetitions predicts better long-term memory—has received support from at least two fMRI studies investigating representational similarity in the HC. That is, representational variability (or instability) in hippocampal patterns across repetitions promotes subsequent memory (Karlsson Wirebring et al., 2015) and memory updating (Speer et al., 2021). These results differ markedly from similar investigations in the neocortex, where it has been found instead that stable patterns across repetitions of the same stimuli promote memory (Xue et al., 2010). One proposed resolution to these disparate findings is that cortical fidelity is required to represent the content of memories, and thereby ensures proper reactivation of old traces, but variability within the HC promotes greater subsequent updating (Karlsson Wirebring et al., 2015). This interpretation fits well with our account, which requires stable inputs about the informational content of cues and targets from elsewhere in the cortex into the EC–HC system but also needs variability from temporal context drift to drive optimal learning via EDL. One potentially conflicting result comes from an important recent study showing that the extent to which prior encoding patterns of picture stimuli are reinstated in the HC during re-encoding days to months later predicts subsequent temporal memory for when the picture was first shown (Zou et al., 2023). It is unclear whether this result, whereby pattern stability rather than variability in HC benefits memory, is specific to the temporal memory task, whether there is some other form of undetected neural variability that serves as a strengthening mechanism, or whether it is problematic for the proposed account.

Our results also have relevance for learning within specific HC pathways. The trisynaptic route involving DG formed associations that continued to change and rely on temporal context with drift between training epochs (Pereira et al., 2007), which allowed for multiple traces of a memory that otherwise had identical content like the cue and target (Guzman et al., 2021). This finding aligns with the idea that the drifting excitability shown in DG neurons may be responsible for time encoding (Aimone et al., 2006, 2009). Intriguingly, we found the EC → CA3 pathway to be critical for the drift-resistant memory component. A prior computational model suggested that this pathway is especially important for generalizing over a variable set of examples belonging to the same visual category at retrieval (Kowadlo et al., 2019). Like our results, Kowadlo et al. (2019) generalization effects were similarly much stronger for this pathway than for the CA3 → CA3 recurrent collaterals, suggesting a prominent role in the EC → CA3 pathway at retrieval. In accordance with these results, our effects can be thought of as creating representations that generalize across time, avoiding the pitfalls of memory that occur without any temporal context support.

Due to the profound long-term benefits of spacing, it is natural to ask whether spacing promotes systems consolidation or the strengthening of the more relatively long-lasting synapses in the neocortex and the gradual ability to recall memories independent of the HC (Attardo et al., 2015; Carpenter, 2020; C. D. Smith & Scarf, 2017). However, current evidence on this is unclear. To our knowledge, there is no current neuroimaging evidence that suggests that spacing decreases hippocampal involvement. Rather, following spaced versus massed learning, the HC shows increased activity (Li & Yang, 2020; Nonaka et al., 2017) or greater connectivity with neocortex (Ezzyat et al., 2018) at retrieval. Evidence from amnesics on the necessity of the HC for spacing effects is also inconclusive. Amnesics show no spacing benefits for recollection, a process that requires the HC, but intact spacing benefits for familiarity (Verfaellie et al., 2008), which can rely on neocortical sources (Yonelinas, 2002). Developmental amnesics do show spacing benefits in free recall, but their hippocampal size was only diminished by 30%, indicating that benefits could result from residual hippocampal tissue (Green et al., 2014). Our results suggest that spacing benefits could be explained without systems consolidation. However, we hope to test the impact of the cortex in later models, especially given evidence that a related phenomenon known as the testing effect—whereby long-term memory benefits more from retrieval than restudy—results in greater neocortical involvement (Antony et al., 2017; Ferreira et al., 2019; Himmer et al., 2019; Van den Broek et al., 2013). One particularly interesting speculation along these lines is that systems consolidation may occur as EDL in multiple stages from HC to cortex, with each faster learning region training each slower learning region (Irish & Vatansever, 2020; Remme et al., 2021; Schapiro et al., 2017).

One novel aspect of our results for memory consolidation research is that a pathway *within* the HC (EC → CA3) produced decontextualization. Here, we defined decontextualization as the invariance to supporting contextual input. In our initial simulations, this invariance arose to changing temporal contexts. Following findings from others that decontextualization can also arise due to invariance of learning environment (Imundo et al., 2021; S. M.

Smith et al., 1978; S. M. Smith & Rothkopf, 1984), semantic context (Beheydt, 1987), and learning task (S. M. Smith & Handy, 2014, 2016), we later simulated how variable learning contexts apart from temporal drift led to similar benefits over constant learning contexts. Although, to our knowledge, prior paradigms have not investigated the neural locus of how decontextualization develops, a related idea called semanticization from a prominent consolidation theory (Winocur et al., 2010; Yassa & Reagh, 2013) occurs when memories lose their contextual details and retain only their central (gist-like) aspects. This theory accounts for findings that rodents with HC damage are more likely to generalize fear memories (Winocur et al., 2010), that fear memories become more generalized over time when the HC aspect of the memory trace becomes less activated (Kitamura et al., 2017; Wiltgen & Silva, 2007), and that HC amnesiac patients often appear normal when asked gist-based questions but impaired when pressed for contextual details (Nadel & Moscovitch, 1997). Therefore, semanticization is only explicitly stated to occur within the neocortex, whereas the HC is stated to support the retrieval of contextual details. Our findings showing decontextualization in HC do not oppose such a neocortical mechanism but instead predict that such a process could initiate the formation of decontextualized traces within the HC itself (Y. Norman et al., 2021; Quiroga et al., 2005).

Limitations

One limitation of HipSTeR is that drift-driven contextual change is simulated as a slow, passive process that is constant per unit time. However, context can also shift more rapidly (DuBrow et al., 2017), which often occurs with sudden input changes or shifts in perceived events (Antony et al., 2021; Baldassano et al., 2017; Bright et al., 2020; Brunec et al., 2018, 2020; Clewett et al., 2019; Cohn-Sheehy et al., 2021; DuBrow & Davachi, 2013, 2014, 2016; Griffiths & Fuentemilla, 2020; Lu et al., 2020; Michelmann et al., 2021; Rouhani et al., 2020; Sellevoll et al., 2023; Wen & Egner, 2022; Zacks et al., 2007). In other words, in addition to slow drifts, there are faster shifts, which allow setting up "walls" between dissimilar situational contexts nearby in time and "bridges" to similar situational contexts apart in time (Clewett et al., 2019; Cohn-Sheehy et al., 2021). Moreover, these sudden event changes can lead to more rapid forgetting of information that is more likely to rely on properly instating the temporal context (Delaney et al., 2010; El-Kalliny et al., 2019; Horner et al., 2016; Radvansky et al., 2011), suggesting that such shifts accelerate the rate of contextual change away from the encoding context. Intriguingly, these shifts have been instantiated in sudden changes in EC (Bright et al., 2020; Tsao et al., 2018; Umbach et al., 2020) and HC activity (Griffiths & Fuentemilla, 2020), and the same EC neurons can show mixed, item-specific patterns of activity that shift with given inputs and also drift thereafter (Bright et al., 2020). Although our final environmental/pictorial decontextualization experiments begin to model such faster shifts, we hope to explore these functional activities in later versions of the model.

Another limitation is that other EC–HC subregions contribute to memory. Specifically, we have not separated LEC, which has neurons showing drifting properties (Tsao et al., 2018), from medial entorhinal cortex. Additionally, when the HC subfields have been compared within the study in a representational drift setting, the subfield most intimately linked with changes in temporal context is CA2 (Mankin et al., 2015). CA2 and CA3 share many similarities yet also have critical differences (Dudek et al., 2016). Here, we worked under the assumption that the EC and CA3 layers in HipSTeR combined aspects that would encompass properties of the dissociable regions. This could elucidate some aspects of the effects of CA3 in time, such as the fact that representational drift in CA1 is slower than CA3 in our model (Figure 2D) but is faster than CA3 in empirical data (Mankin et al., 2015). This could be reconciled by the fact that drift is faster in CA2 than any of the other CA regions, and these CA2 properties may be included in our CA3.

A final limitation of our model is that we did not capture some of the spacing effects produced by expanding spacing schedules. Instances in which expanding spacing is superior to equal spacing seem limited to cases in which feedback is not given after tests (Cull et al., 1996; Landauer & Bjork, 1978). In these training regimes, it is fairly unlikely that memories will be recovered once lost, so rehearsing after an early interval after learning keeps the memory "in the running," so to speak, whereas a slightly longer initial interval (with equal spacing) may result in a higher percentage of memories being lost and unrecoverable. In instances when feedback is given or when restudy is used rather than retrieval, there is often no benefit of expanding spacing over equal spacing (Carpenter & DeLosh, 2005; Cull, 2000; Cull et al., 1996; Karpicke & Roediger, 2010; Küpper-Tetzel et al., 2014; Landauer & Bjork, 1978). These instances are more analogous to our model, as the target information is always available during training epochs. However, expanding performance in our model contradicted the behavioral data by being inferior to an equal schedule and not superior to a contracting schedule. There do not seem to be convincing theoretical explanations for this effect, though Küpper-Tetzel et al. (2014) have speculated that the impact of later ISIs may have a stronger influence on memory performance than earlier ISIs. Perhaps later versions of the model could attempt to employ related mechanisms.

Conclusions and Broader Relevance

Given a lifetime of learning, our goal should be to optimize knowledge that can be retained over long periods of time. In this regard, the practical difficulties with conducting and publishing well-controlled research studies with long delays have likely biased the literature toward what can be assessed on relatively short time scales. Nevertheless, it is difficult to overstate the importance of spacing in long-term learning (Bahrick, 1979; Rawson & Dunlosky, 2011): even the well-known importance of testing for learning (e.g., Rowland, 2014) cannot overcome poor temporal generalization when tests are conducted at short lags, as practicing paired associates up to 10 times with 1-minute lags can still produce floor (<5% correct) performance a week later (Pyc & Rawson, 2009). Ultimately, a memory that remains context-dependent (e.g., Abernethy, 1940; Imundo et al., 2021; Parker et al., 2007; S. M. Smith et al., 1978; S. M. Smith & Rothkopf, 1984) may not be ideal for learning because it continues to rely on the "crutch" of context—either the presence of contextual inputs or the mental recovery of a context—for successful memory retrieval (S. M. Smith & Handy, 2014). Efforts to render memories context-independent when the context is an incidental or irrelevant aspect of the memory will ultimately benefit lifelong learning.

Recent attempts to overcome catastrophic interference and other kinds of ordering effects, such as how prior learning generalizes to new tasks, have led the artificial intelligence field to the problem of continual learning (e.g., Flesch et al., 2018; Hadsell et al., 2020; Masse et al., 2018; van de Ven et al., 2020). We hope our efforts will demonstrate the critical importance of accounting for temporal context in artificial intelligence problems. In particular, the manner in which drift between training trials—featuring different input patterns—actually improved performance in HipSTeR seems particularly important in addressing these problems.

Though representations in some regions are relatively stable over time (Dhawale et al., 2017; Katlowitz et al., 2018; McGuire et al., 2022; Pérez-Ortega et al., 2021), drift occurs in many regions beyond the medial temporal lobe (Deitch et al., 2021; Driscoll et al., 2017; Hyman et al., 2012; Margolis et al., 2012). Notably, spacing effects have been observed in numerous other learning domains, including motor-skill learning (e.g., Baddeley & Longman, 1978; T. D. Lee & Genovese, 1988; Shea et al., 2000), the acquisition of math skills (e.g., Rohrer, 2009), reading (e.g., Krug et al., 1990), classical conditioning (e.g., Rohrer, 2009) and extinction (e.g., Bernal-Gamboa et al., 2018; Rowe & Craske, 1998), and similar EDL mechanisms may be at play in the systems underlying these skills. Intriguingly, computational models have suggested multi-tiered learning in other domains like motor learning (Murray & Escola, 2020), suggesting parallel mechanisms in strengthening increasingly long-lasting memory traces in these systems. Additionally, there are known biological ramifications of spacing, such that spaced rather than massed inductions of long-term potentiation result in a slower decline in long-term potentiation-induced changes (Jiang et al., 2016; Josselyn et al., 2001; Kramár et al., 2012; Scharf et al., 2002; Smolen et al., 2016). Differences in these properties across brain regions may play a critical role in how spacing and drift affect learning.

Finally, our investigations have implications for those with impairments within the EC–HC system. Anterolateral EC hypoactivity occurs in healthy aging and has been linked with memory deficits (Reagh et al., 2018), and anterolateral EC is one of the first regions to be affected by Alzheimer's Disease (e.g., Braak & Braak, 1995). Other long-term memory abnormalities linked with temporal lobe impairments, including accelerated long-term forgetting, which involves rapid forgetting after initially intact long-term memories (>30 min; Elliott et al., 2014), could ultimately rely on mechanisms uncovered here.

References

Abernethy, E. M. (1940). The effect of changed environmental conditions upon the results of college examinations. *The Journal of Psychology*, *10*(2), 293–301. https://doi.org/10.1080/00223980.1940.9917005

Aghajan, Z. M., Kreiman, G., & Fried, I. (2023). Minute-scale periodicity of neuronal firing in the human entorhinal cortex. *Cell Reports*, *42*(11), Article 113271. https://doi.org/10.1016/j.celrep.2023.113271

Aimone, J. B., Wiles, J., & Gage, F. H. (2006). Potential role for adult neurogenesis in the encoding of time in new memories. *Nature Neuroscience*, *9*(6), 723–727. https://doi.org/10.1038/nn1707

Aimone, J. B., Wiles, J., & Gage, F. H. (2009). Computational influence of adult neurogenesis on memory encoding. *Neuron*, *61*(2), 187–202. https://doi.org/10.1016/j.neuron.2008.11.026

Anderson, J. R., & Milson, R. (1989). Human memory: An adaptive perspective. *Psychological Review*, *96*(4), 703–719. https://doi.org/10.1037/0033-295X.96.4.703

Anderson, J. R., & Schooler, L. J. (1991). Reflections of the environment in memory. *Psychological Science*, *2*(6), 396–408. https://doi.org/10.1111/j.1467-9280.1991.tb00174.x

Antony, J. W., Ferreira, C. S., Norman, K. A., & Wimber, M. (2017). Retrieval as a fast route to memory consolidation. *Trends in Cognitive Sciences*, *21*(8), 573–576. https://doi.org/10.1016/j.tics.2017.05.001

Antony, J. W., Hartshorne, T. H., Pomeroy, K., Gureckis, T. M., Hasson, U., McDougle, S. D., & Norman, K. A. (2021). Behavioral, physiological, and neural signatures of surprise during naturalistic sports viewing. *Neuron*, *109*(2), 377–390. https://doi.org/10.1016/j.neuron.2020.10.029

Arzy, S., Adi-Japha, E., & Blanke, O. (2009). The mental time line: An analogue of the mental number line in the mapping of life events. *Consciousness and Cognition*, *18*(3), 781–785. https://doi.org/10.1016/j.concog.2009.05.007

Attardo, A., Fitzgerald, J. E., & Schnitzer, M. J. (2015). Impermanence of dendritic spines in live adult CA1 hippocampus. *Nature*, *523*(7562), 592–596. https://doi.org/10.1038/nature14467

Baddeley, D., & Longman, D. J. (1978). The influence of length and frequency of training session on the rate of learning to type. *Ergonomics*, *21*(8), 627–635. https://doi.org/10.1080/00140137808931764

Bahrick, H. P. (1979). Maintenance of knowledge: Questions about memory we forgot to ask. *Journal of Experimental Psychology: General*, *108*(3), 296–308. https://doi.org/10.1037/0096-3445.108.3.296

Bahrick, H. P., Bahrick, L. E., Bahrick, A. S., & Bahrick, P. E. (1993). Maintenance of foreign language vocabulary and the spacing effect. *Psychological Science*, *4*(5), 316–321. https://doi.org/10.1111/j.1467-9280.1993.tb00571.x

Bakker, A., Kirwan, C. B., Miller, M., & Stark, C. E. L. (2008). Pattern separation in the human hippocampal CA3 and dentate gyrus. *Science*, *319*(5870), 1640–1642. https://doi.org/10.1126/science.1152882

Baldassano, C., Chen, J., Zadbood, A., Pillow, J. W., Hasson, U., & Norman, K. A. (2017). Discovering event structure in continuous narrative perception and memory. *Neuron*, *95*(3), 709–721.e5. https://doi.org/10.1016/j.neuron.2017.06.041

Balota, D. A., Duchek, J. M., & Logan, J. M. (2007). Is expanded retrieval practice a superior form of spaced retrieval? A critical review of the extant literature. In J. S. Nairne (Ed.), *The foundations of remembering: Essays in honor of Henry L. Roediger, III* (pp. 83–105). Psychology Press.

Balota, D. A., Duchek, J. M., & Paullin, R. (1989). Age-related differences in the impact of spacing, lag, and retention interval. *Psychology and Aging*, *4*(1), 3–9. https://doi.org/10.1037/0882-7974.4.1.3

Barnett, A. J., O'Neil, E. B., Watson, H. C., & Lee, A. C. (2014). The human hippocampus is sensitive to the durations of events and intervals within a sequence. *Neuropsychologia*, *64*, 1–12. https://doi.org/10.1016/j.neuropsychologia.2014.09.011

Begg, I., & Green, C. (1988). Repetition and trace interaction: Superadditivity. *Memory & Cognition*, *16*(3), 232–242. https://doi.org/10.3758/BF03197756

Beheydt, L. (1987). The semanticization of vocabulary in foreign language learning. *System*, *15*(1), 55–67. https://doi.org/10.1016/0346-251X(87)90048-0

Bellmund, J. L., Deuker, L., & Doeller, C. F. (2019). Mapping sequence structure in the human lateral entorhinal cortex. *eLife*, *8*, Article e45333. https://doi.org/10.7554/eLife.45333

Benjamin, A. S., & Tullis, J. (2010). What makes distributed practice effective? *Cognitive Psychology*, *61*(3), 228–247. https://doi.org/10.1016/j.cogpsych.2010.05.004

Bernal-Gamboa, R., Gámez, A. M., & Nieto, J. (2018). Spacing extinction sessions as a behavioral technique for preventing relapse in an animal model of voluntary actions. *Behavioural Processes*, *151*, 54–61. https://doi.org/10.1016/j.beproc.2018.01.021

Bjork, R. A. (2011). On the symbiosis of learning, remembering, and forgetting. *Successful remembering and successful forgetting: A Festschrift in honor of Robert A. Bjork* (pp. 1–22). Psychology Press.

Bjork, R. A., & Bjork, E. (1992). A new theory of disuse and an old theory of stimulus fluctuation. In A. Healy, S. Kosslyn, & R. Shiffrin (Eds.), *Learning processes to cognitive processes: Essays in Honor of William K. Estes* (Vol. 2, pp. 35–67). Hillsdale.

Bladon, J. H., Sheehan, D. J., De Freitas, C. S., & Howard, M. W. (2019). In a temporally segmented experience hippocampal neurons represent temporally drifting context but not discrete segments. *Journal of Neuroscience*, *39*(35), 6936–6952. https://doi.org/10.1523/JNEUROSCI.1420-18.2019

Bousfield, W. A. (1953). The occurrence of clustering in the recall of randomly arranged associates. *The Journal of General Psychology*, *49*(2), 229–40. https://doi.org/10.1080/00221309.1953.9710088

Bouton, M. E. (1993). Context, time, and memory retrieval in the interference paradigms of Pavlovian learning. *Psychological Bulletin*, *114*(1), 80–99. https://doi.org/10.1037/0033-2909.114.1.80

Bower, G. (1967). A multicomponent theory of the memory trace. *The psychology of learning and motivation: Advances in research and theory* (pp. 229–325). Academic Press.

Braak, H., & Braak, E. (1995). Staging of Alzheimer's Disease-related neurofibrillary changes. *Neurobiology of Aging*, *16*(3), 31–33. https://doi.org/10.1016/0197-4580(95)00021-6

Bradburn, N. M., Rips, L. J., & Shevell, S. K. (1987). Answering autobiographical questions: The impact of Memory and Inference on Surveys. *Science*, *236*(4798), 157–161. https://doi.org/10.1126/science.3563494

Brea, J., Urbanczik, R., & Senn, W. (2014). A normative theory of forgetting: Lessons from the fruit fly. *PLOS Computational Biology*, *10*(6), Article e1003640. https://doi.org/10.1371/journal.pcbi.1003640

Brietzke, S., & Meyer, M. L. (2021). Temporal self-compression: Behavioral and neural evidence that past and future selves are compressed as they move away from the present. *Proceedings of the National Academy of Sciences of the United States of America*, *118*(49), Article e2101403118. https://doi.org/10.1073/pnas.2101403118

Bright, I. M., Meister, M. L., Cruzado, N. A., Tiganj, Z., Buffalo, E. A., & Howard, M. W. (2020). A temporal record of the past with a spectrum of time constants in the monkey entorhinal cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(33), 20274–20283. https://doi.org/10.1073/pnas.1917197117

Brod, G. (2021). Predicting as a learning strategy. *Psychonomic Bulletin and Review*, *28*(6), 1839–1847. https://doi.org/10.3758/s13423-021-01904-1

Brown, G. D., Neath, I., & Chater, N. (2007). A temporal ratio model of memory. *Psychological Review*, *114*(3), 539–576. https://doi.org/10.1037/0033-295X.114.3.539

Brozinsky, C. J., Yonelinas, A. P., Kroll, N. E., & Ranganath, C. (2005). Lag-sensitive repetition suppression effects in the anterior parahippocampal gyrus. *Hippocampus*, *15*, 557–561. https://doi.org/10.1002/hipo.20087

Brunec, I. K., Moscovitch, M., & Barense, M. D. (2018). Boundaries shape cognitive representations of spaces and events. *Trends in Cognitive Sciences*, *22*(7), 637–650. https://doi.org/10.1016/j.tics.2018.03.013

Brunec, I. K., Ozubko, J. D., Ander, T., Guo, R., Moscovitch, M., & Barense, M. D. (2020). Turns during navigation act as boundaries that enhance spatial memory and expand time estimation. *Neuropsychologia*, *141*, Article 107437. https://doi.org/10.1016/j.neuropsychologia.2020.107437

Buonomano, D. V., & Merzenich, M. M. (1995). Temporal information transformed into a spatial code by a neural network with realistic properties. *Science*, *267*(5200), 1028–1030. https://doi.org/10.1126/science.7863330

Burt, C. D. (1992). Retrieval characteristics of autobiographical memories: Event and date information. *Applied Cognitive Psychology*, *6*(5), 389–404. https://doi.org/10.1002/acp.2350060504

Butler, A. C., Black-Maier, A. C., Raley, N. D., & Marsh, E. J. (2017). Retrieving and applying knowledge to different examples promotes transfer of learning. *Journal of Experimental Psychology: Applied*, *23*(4), 433–446. https://doi.org/10.1037/xap0000142

Cai, D. J., Aharoni, D., Shuman, T., Shobe, J., Biane, J., Song, W., Wei, B., Veshkini, M., La-Vu, M., Lou, J., Flores, S. E., Kim, I., Sano, Y., Zhou, M., Baumgaertel, K., Lavi, A., Kamata, M., Tuszynski, M., Mayford, M., … Silva, A. J. (2016). A shared neural ensemble links distinct contextual memories encoded close in time. *Nature*, *534*(7605), 115–118. https://doi.org/10.1038/nature17955

Cao, R., Bladon, J. H., Charczynski, S. J., Hasselmo, M. E., & Howard, M. W. (2022). Internally generated time in the rodent hippocampus is logarithmically compressed. *eLife*, *11*, Article e75353. https://doi.org/10.7554/eLife.75353

Carpenter, S. K. (2020). Distributed practice or spacing effect. *Oxford research encyclopedia of education*. Oxford University Press. https://doi.org/10.1093/acrefore/9780190264093.013.859

Carpenter, S. K., & DeLosh, E. L. (2005). Application of the testing and spacing effects to name learning. *Applied Cognitive Psychology*, *19*(5), 619–36. https://doi.org/10.1002/acp.1101

Cepeda, N. J., Coburn, N., Rohrer, D., Wixted, J. T., Mozer, M. C., & Pashler, H. (2009). Optimizing distributed practice theoretical analysis and practical implications. *Experimental Psychology*, *56*(4), 236–246. https://doi.org/10.1027/1618-3169.56.4.236

Cepeda, N. J., Pashler, H., Vul, E., Wixted, J. T., & Rohrer, D. (2006). Distributed practice in verbal recall tasks: A review and quantitative synthesis. *Psychological Bulletin*, *132*(3), 354–380. https://doi.org/10.1037/0033-2909.132.3.354

Cepeda, N. J., Vul, E., Rohrer, D., Wixted, J. T., & Pashler, H. (2008). Spacing effects in learning. *Psychological Science*, *19*(11), 1095–102. https://doi.org/10.1111/j.1467-9280.2008.02209.x

Chen, J., Honey, C. J., Simony, E., Arcaro, M. J., Norman, K. A., & Hasson, U. (2016). Accessing real-life episodic information from minutes versus hours earlier modulates hippocampal and high-order cortical dynamics. *Cerebral Cortex*, *26*(8), 3428–3441. https://doi.org/10.1093/cercor/bhv155

Clewett, D., Dubrow, S., & Davachi, L. (2019). Transcending time in the brain: How event memories are constructed from experience. *Hippocampus*, *29*(3), 162–83. https://doi.org/10.1002/hipo.23074

Clopath, C., Bonhoeffer, T., Hübener, M., & Rose, T. (2017). Variance and invariance of neuronal long-term representations. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *372*(1715), Article 20160161. https://doi.org/10.1098/rstb.2016.0161

Cohn-Sheehy, B. I., Delarazan, A. I., Reagh, Z. M., Crivelli-Decker, J. E., Kim, K., Barnett, A. J., Zacks, J. M., & Ranganath, C. (2021). Building bridges: The hippocampus constructs narrative memories across distant events. *Current Biology*, *31*(22), 4935–4945.e7. https://doi.org/10.1016/j.cub.2021.09.013

Colgin, L. L., Moser, E. I., & Moser, M.-B. (2008). Understanding memory through hippocampal remapping. *Trends in Neurosciences*, *31*(9), 469–77. https://doi.org/10.1016/j.tins.2008.06.008

Coulter, D. A., & Carlson, G. C. (2007). Functional regulation of the dentate gyrus by GABA-mediated inhibition. *Progress in Brain Research*, *163*, 235–812. https://doi.org/10.1016/S0079-6123(07)63014-3

Cox, W. R., Dobbelaar, S., Meeter, M., Kindt, M., & van Ast, V. A. (2021). Episodic memory enhancement versus impairment is determined by contextual similarity across events. *Proceedings of the National Academy of Sciences of the United States of America*, *118*(48), Article e2101509118. https://doi.org/10.1073/pnas.2101509118

Crowder, R. G. (1976). *Principles of learning and memory: Classic edition*. Erlbaum.

Cull, W. L. (2000). Untangling the benefits of multiple study opportunities and repeated testing for cued recall. *Applied Cognitive Psychology*, *14*(3), 215–35. https://doi.org/10.1002/(SICI)1099-0720(200005/06)14:3<215::AID-ACP640>3.0.CO;2-1

Cull, W. L., Shaughnessy, J. J., & Zechmeister, E. B. (1996). Expanding understanding of the expanding-pattern-of-retrieval mnemonic: Toward confidence in applicability. *Journal of Experimental Psychology: Applied*, *2*(4), 365–378. https://doi.org/10.1037/1076-898X.2.4.365

Davachi, L. (2006). Item, context and relational episodic encoding in humans. *Current Opinion in Neurobiology*, *16*(6), 693–700. https://doi.org/10.1016/j.conb.2006.10.012

Dehaene, S. (2003). The neural basis of the Weber-Fechner law: A logarithmic mental number line. *Trends in Cognitive Sciences*, *7*(4), 145–147. https://doi.org/10.1016/S1364-6613(03)00055-X

Deitch, D., Rubin, A., & Ziv, Y. (2021). Representational drift in the mouse visual cortex. *Current Biology*, *31*(19), 4327–4339.e6. https://doi.org/10.1016/j.cub.2021.07.062

Delamare, G., Zaki, Y., Cai, D. J., & Clopath, C. (2023). *Drift of neural ensembles driven by slow fluctuations of intrinsic excitability*. bioRxiv. https://doi.org/10.1101/2023.03.16.532958

Delaney, P. F., Sahakyan, L., Kelley, C. M., & Zimmerman, C. A. (2010). Remembering to forget: The amnesic effect of daydreaming. *Psychological Science*, *21*(7), 1036–1042. https://doi.org/10.1177/0956797610374739

Dester, A. M., Lazarus, L. T. T., Uitvlugt, M. G., & Healey, M. K. (2021). A test of retrieved context theory: Dynamics of recall after incidental encoding. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *47*(8), 1264–1287. https://doi.org/10.1037/xlm0001001

Devalle, F., & Roxin, A. (2022). *Network mechanisms underlying representational drift in area CA1 of hippocampus*. bioRxiv. https://doi.org/10.1101/2022.11.10.515946

Dhawale, A. K., Poddar, R., Wolff, S. B., Normand, V. A., Kopelowitz, E., & Ölveczky, B. P. (2017). Automated long-term recording and analysis of neural activity in behaving animals. *eLife*, *6*, Article e27702. https://doi.org/10.7554/eLife.27702

Dimsdale-Zucker, H. R., Montchal, M. E., Reagh, Z. M., Wang, S.-F., Libby, L. A., & Ranganath, C. (2022). Representations of complex contexts: A role for hippocampus. *Journal of Cognitive Neuroscience*, *35*(1), 90–110. https://doi.org/10.1162/jocn_a_01919

Driscoll, L. N., Pettit, N. L., Minderer, M., Chettih, S. N., & Harvey, C. D. (2017). Dynamic reorganization of neuronal activity patterns in parietal cortex. *Cell*, *170*(5), 986–999.e16. https://doi.org/10.1016/j.cell.2017.07.021

Driskell, J. E., Willis, R. P., & Copper, C. (1992). Effect of overlearning on retention. *Journal of Applied Psychology*, *77*(5), 615–622. https://doi.org/10.1037/0021-9010.77.5.615

DuBrow, S., & Davachi, L. (2013). The influence of context boundaries on memory for the sequential order of events. *Journal of Experimental Psychology: General*, *142*(4), 1277–1286. https://doi.org/10.1037/a0034024

DuBrow, S., & Davachi, L. (2014). Temporal memory is shaped by encoding stability and intervening item reactivation. *Journal of Neuroscience*, *34*(42), 13998–14005. https://doi.org/10.1523/JNEUROSCI.2535-14.2014

DuBrow, S., & Davachi, L. (2016). Temporal binding within and across events. *Neurobiology of Learning and Memory*, *134*, 107–114. https://doi.org/10.1016/j.nlm.2016.07.011

DuBrow, S., Rouhani, N., Niv, Y., & Norman, K. A. (2017). Does mental context drift or shift? *Current Opinion in Behavioral Sciences*, *17*, 141–146. https://doi.org/10.1016/j.cobeha.2017.08.003

Dudek, S. M., Alexander, G. M., & Farris, S. (2016). Rediscovering area CA2: Unique properties and functions. *Nature Reviews Neuroscience*, *17*(2), 89–102. https://doi.org/10.1038/nrn.2015.22

Dusek, J. A., & Eichenbaum, H. (1997). The hippocampus and memory for orderly stimulus relations. *Proceedings of the National Academy of Sciences of the United States of America*, *94*(13), 7109–7114. https://doi.org/10.1073/pnas.94.13.7109

Ebbinghaus, H. (1885). *Memory: A contribution to experimental psychology*. Columbia University Press.

Egorov, A. V., Hamam, B. N., Fransén, E., Hasselmo, M. E., & Alonso, A. A. (2002). Graded persistent activity in entorhinal cortex neurons. *Nature*, *420*(6912), 173–178. https://doi.org/10.1038/nature01171

Eichenbaum, H., Yonelinas, A. P., & Ranganath, C. (2007). The medial temporal lobe and recognition memory. *Annual Review of Neuroscience*, *30*(1), 123–152. https://doi.org/10.1146/annurev.neuro.30.051606.094328

El-Kalliny, M. M., Wittig, J. H., Sheehan, T. C., Sreekumar, V., Inati, S. K., & Zaghloul, K. A. (2019). Changing temporal context in human temporal lobe promotes memory of distinct episodes. *Nature Communications*, *10*(1), Article 203. https://doi.org/10.1038/s41467-018-08189-4

Elliott, G., Isaac, C. L., & Muhlert, N. (2014). Measuring forgetting: A critical review of accelerated long-term forgetting studies. *Cortex*, *54*(1), 16–32. https://doi.org/10.1016/j.cortex.2014.02.001

Elman, J. L., & Zipser, D. (1988). Learning the hidden structure of speech. *Journal of the Acoustical Society of America*, *83*(4), 1615–1626. https://doi.org/10.1121/1.395916

Estes, W. K. (1955). Statistical theory of distributional phenomena in learning. *Psychological Review*, *62*(5), Article 369. https://doi.org/10.1037/h0046888

Ezzyat, Y., & Davachi, L. (2014). Similarity breeds proximity: Pattern similarity within and across contexts is related to later mnemonic judgments of temporal proximity. *Neuron*, *81*(5), 1179–1189. https://doi.org/10.1016/j.neuron.2014.01.042

Ezzyat, Y., Inhoff, M. C., & Davachi, L. (2018). Differentiation of human medial prefrontal cortex activity underlies long-term resistance to forgetting in memory. *Journal of Neuroscience*, *38*(48), 10244–54. https://doi.org/10.1523/JNEUROSCI.2290-17.2018

Fenton, A. A., Kao, H. Y., Neymotin, S. A., Olypher, A., Vayntrub, Y., Lytton, W. W., & Ludvig, N. (2008). Unmasking the CA1 ensemble place code by exposures to small and large environments: More place cells and multiple, irregularly arranged, and expanded place fields in the larger space. *Journal of Neuroscience*, *28*(44), 11250–11262. https://doi.org/10.1523/JNEUROSCI.2862-08.2008

Ferreira, C. S., Charest, I., & Wimber, M. (2019). Retrieval aids the creation of a generalised memory trace and strengthens episode-unique information. *NeuroImage*, *201*, Article 115996. https://doi.org/10.1016/j.neuroimage.2019.07.009

Flesch, T., Balaguer, J., Dekker, R., Nili, H., & Summerfield, C. (2018). Comparing continual task learning in minds and machines. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(44), E10313–E10322. https://doi.org/10.1073/pnas.1800755115

Folkerts, S., Rutishauser, U., & Howard, M. W. (2018). Human episodic memory retrieval is accompanied by a neural contiguity effect. *Journal of Neuroscience*, *38*(17), 4200–4211. https://doi.org/10.1523/JNEUROSCI.2312-17.2018

Fortin, N. J., Agster, K. L., & Eichenbaum, H. B. (2002). Critical role of the hippocampus in memory for sequences of events. *Nature Neuroscience*, *5*(5), 458–462. https://doi.org/10.1038/nn834

Frank, A. C., Huang, S., Zhou, M., Gdalyahu, A., Kastellakis, G., Silva, T. K., Lu, E., Wen, X., Poirazi, P., Trachtenberg, J. T., & Silva, A. J. (2018). Hotspots of dendritic spine turnover facilitate clustered spine addition and learning and memory. *Nature Communications*, *9*(1), Article 422. https://doi.org/10.1038/s41467-017-02751-2

Friedman, W. J. (1993). Memory for the time of past events. *Psychological Bulletin*, *113*(1), 44–66. https://doi.org/10.1037/0033-2909.113.1.44

Gallego, J. A., Perich, M. G., Chowdhury, R. H., Solla, S. A., & Miller, L. E. (2020). Long-term stability of cortical population dynamics underlying consistent behavior. *Nature Neuroscience*, *23*(2), 260–270. https://doi.org/10.1038/s41593-019-0555-4

Gerbier, E., Toppino, T. C., & Koenig, O. (2015). Optimising retention through multiple study opportunities over days: The benefit of an expanding schedule of repetitions. *Memory*, *23*(6), 943–954. https://doi.org/10.1080/09658211.2014.944916

Gershman, S. J., & Niv, Y. (2010). Learning latent structure: Carving nature at its joints. *Current Opinion in Neurobiology*, *20*(2), 251–256. https://doi.org/10.1016/j.conb.2010.02.008

Geva, N., Deitch, D., Rubin, A., & Ziv, Y. (2023). Time and experience differentially affect distinct aspects of hippocampal representational drift. *Neuron*, *111*(15), 2357–2366. https://doi.org/10.1016/j.neuron.2023.05.005

Gibbon, J. (1977). Scalar expectancy theory and Weber's law in animal timing. *Psychological Review*, *84*(3), 279–325. https://doi.org/10.1037/0033-295X.84.3.279

Gibbon, J., Church, R. M., & Meck, W. H. (1984). Scalar timing in memory. *Annals of the New York Academy of Sciences*, *423*(1), 52–77. https://doi.org/10.1111/j.1749-6632.1984.tb23417.x

Glass, A. L. (2009). The effect of distributed questioning with varied examples on exam performance on inference questions. *Educational Psychology*, *29*(7), 831–848. https://doi.org/10.1080/01443410903310674

Glenberg, A. M. (1976). Monotonic and nonmonotonic lag effects in paired-associate and recognition memory paradigms. *Journal of Verbal Learning and Verbal Behavior*, *15*(1), 1–16. https://doi.org/10.1016/S0022-5371(76)90002-5

Glenberg, A. M. (1979). Component-levels theory of the effects of spacing of repetitions on recall and recognition. *Memory & Cognition*, *7*(2), 95–112. https://doi.org/10.3758/BF03197590

Godden, D., & Baddeley, A. (1975). Context-dependent memory in two natural environments: On land and underwater. *British Journal of Psychology*, *66*(3), 325–31. https://doi.org/10.1111/j.2044-8295.1975.tb01468.x

Green, J. L., Weston, T., Wiseheart, M., & Rosenbaum, R. S. (2014). Long-term spacing effect benefits in developmental amnesia: Case experiments in rehabilitation. *Neuropsychology*, *28*(5), 685–694. https://doi.org/10.1037/neu0000070

Greene, R. L. (1989). Spacing effects in memory: Evidence for a two-process account. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*(3), 371–377. https://doi.org/10.1037/0278-7393.15.3.371

Grienberger, C., Magee, J. C., & Duncan, D. (2022). Entorhinal cortex directs learning-related changes in CA1 representations. *Nature*, *611*(7636), 554–562. https://doi.org/10.1038/s41586-022-05378-6

Griffiths, B. J., & Fuentemilla, L. (2020). Event conjunction: How the hippocampus integrates episodic memories across event boundaries. *Hippocampus*, *30*(2), 162–171. https://doi.org/10.1002/hipo.23161

Grossberg, S. (1982). *Studies of mind and brain: Neural principles of learning, perception, development, cognition, and motor control*. Reidell Press.

Grossberg, S., & Schmajuk, N. A. (1989). Neural dynamics of adaptive timing and temporal discrimination during associative learning. *Neural Networks*, *2*(2), 79–102. https://doi.org/10.1016/0893-6080(89)90026-9

Guo, C., Huson, V., Macosko, E. Z., & Regehr, W. G. (2021). Graded heterogeneity of metabotropic signaling underlies a continuum of cell-intrinsic temporal responses in unipolar brush cells. *Nature Communications*, *12*(1), 2–13. https://doi.org/10.1038/s41467-020-20340-8

Guzman, S. J., Schlögl, A., Espinoza, C., Zhang, X., Suter, B. A., & Jonas, P. (2021). How connectivity rules and synaptic properties shape the efficacy of pattern separation in the entorhinal cortex–dentate gyrus–CA3 network. *Nature Computational Science*, *1*(12), 830–842. https://doi.org/10.1038/s43588-021-00157-1

Hadsell, R., Rao, D., Rusu, A. A., & Pascanu, R. (2020). Embracing change: Continual learning in deep neural networks. *Trends in Cognitive Sciences*, *24*(12), 1028–1040. https://doi.org/10.1016/j.tics.2020.09.004

Hainmueller, T., & Bartos, M. (2018). Parallel emergence of stable and dynamic memory engrams in the hippocampus. *Nature*, *558*(7709), 292–296. https://doi.org/10.1038/s41586-018-0191-2

Hasselmo, M. E. (2005). What is the function of hippocampal theta rhythm? Linking behavioral data to phasic properties of field potential and unit recording data. *Hippocampus*, *15*(7), 936–49. https://doi.org/10.1002/hipo.20116

Hasselmo, M. E., & McGaughy, J. (2004). High acetylcholine levels set circuit dynamics for attention and encoding and low acetylcholine levels set dynamics for consolidation. *Progress in Brain Research*, *145*, 207–31. https://doi.org/10.1016/S0079-6123(03)45015-2

Healey, M. K. (2018). Temporal contiguity in incidentally encoded memories. *Journal of Memory and Language*, *102*, 28–40. https://doi.org/10.1016/j.jml.2018.04.003

Healey, M. K., Long, N. M., & Kahana, M. J. (2019). Contiguity in episodic memory. *Psychonomic Bulletin and Review*, *26*(3), 699–720. https://doi.org/10.3758/s13423-018-1537-3

Hebb, D. O. (1949). *The organization of behavior: A neuropsychological theory*. Lawrence Erlbaum Associates.

Henze, D. A., Wittner, L., & Buzsáki, G. (2002). Single granule cells reliably discharge targets in the hippocampal CA3 network in vivo. *Nature Neuroscience*, *5*(8), 790–795. https://doi.org/10.1038/nn887

Heusser, A. C., Ezzyat, Y., Shiff, I., & Davachi, L. (2018). Perceptual boundaries cause mnemonic trade-offs between local boundary processing and across-trial associative binding. *Journal of Experimental Psychology: Learning Memory and Cognition*, *44*(7), 1075–1090. https://doi.org/10.1037/xlm0000503

Heusser, A. C., Fitzpatrick, P. C., & Manning, J. R. (2021). Geometric models reveal behavioral and neural signatures of transforming naturalistic experiences into episodic memories. *Nature Human Behaviour*, *5*(7), 905–919. https://doi.org/10.1038/s41562-021-01051-6

Himmer, L., Schönauer, M., Heib, D. P. J., Schabus, M., & Gais, S. (2019). Rehearsal initiates systems memory consolidation, sleep makes it last. *Science Advances*, *5*(4), Article eaav1695. https://doi.org/10.1126/sciadv.aav1695

Hinton, G. E., & van Camp, D. (1993). *Keeping the neural networks simple by minimizing the description length of the weights* [Conference session]. Proceedings of the Sixth Annual Conference on Computational Learning Theory.

Hintzman, D. L. (1986). "Schema abstraction" in a multiple-trace memory model. *Psychological Review*, *93*(4), 411–428. https://doi.org/10.1037/0033-295X.93.4.411

Hintzman, D. L. (2016). Is memory organized by temporal contiguity? *Memory and Cognition*, *44*(3), 365–375. https://doi.org/10.3758/s13421-015-0573-8

Horner, A. J., Bisby, J. A., Wang, A., Bogus, K., & Burgess, N. (2016). The role of spatial boundaries in shaping long-term event representations. *Cognition*, *154*, 151–164. https://doi.org/10.1016/j.cognition.2016.05.013

Howard, M. W. (2018). Memory as perception of the past: Compressed time in mind and brain. *Trends in Cognitive Sciences*, *22*(2), 124–136. https://doi.org/10.1016/j.tics.2017.11.004

Howard, M. W., & Kahana, M. J. (2002). A distributed representation of temporal context. *Journal of Mathematical Psychology*, *46*(3), 269–99. https://doi.org/10.1006/jmps.2001.1388

Howard, M. W., Viskontas, I. V., Shankar, K. H., & Fried, I. (2012). Ensembles of human MTL neurons "jump back in time" in response to a repeated stimulus. *Hippocampus*, *22*(9), 1833–1847. https://doi.org/10.1002/hipo.22018

Howard, M. W., Youker, T. E., & Venkatadass, V. S. (2008). The persistence of memory: Contiguity effects across hundreds of seconds. *Psychonomic Bulletin & Review*, *15*(1), 58–63. https://doi.org/10.3758/PBR.15.1.58

Hsieh, L.-T., Gruber, M. J., Jenkins, L. J., & Ranganath, C. (2014). Hippocampal activity patterns carry information about objects in temporal context. *Neuron*, *81*(5), 1165–78. https://doi.org/10.1016/j.neuron.2014.01.015

Huff, M. J., & Bodner, G. E. (2014). All varieties of encoding variability are not created equal: Separating variable processing from variable tasks. *Journal of Memory and Language*, *73*(1), 43–58. https://doi.org/10.1016/j.jml.2014.02.004

Hull, C. L. (1943). *The principles of behavior*. Appleton-Century-Crofts.

Hyman, J. M., Ma, L., Balaguer-Ballester, E., Durstewitz, D., & Seamans, J. K. (2012). Contextual encoding by ensembles of medial prefrontal cortex neurons. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(13), 5086–5091. https://doi.org/10.1073/pnas.1114415109

Imundo, M. N., Pan, S. C., Bjork, E. L., & Bjork, R. A. (2021). Where and how to learn: The interactive benefits of contextual variation, restudying, and retrieval practice for learning. *Quarterly Journal of Experimental Psychology*, *74*(3), 413–424. https://doi.org/10.1177/1747021820968483

Irish, M., & Vatansever, D. (2020). Rethinking the episodic-semantic distinction from a gradient perspective. *Current Opinion in Behavioral Sciences*, *32*, 43–49. https://doi.org/10.1016/j.cobeha.2020.01.016

Jacques, B. G., Tiganj, Z., Sarkar, A., Howard, M., & Sederberg, P. (2022). *A deep convolutional neural network that is invariant to time rescaling* [Conference session]. International Conference on Machine Learning.

Jenkins, L. J., & Ranganath, C. (2010). Prefrontal and medial temporal lobe activity at encoding predicts temporal context memory. *Journal of Neuroscience*, *30*(46), 15558–15565. https://doi.org/10.1523/JNEUROSCI.1337-10.2010

Jeunehomme, O., & D'Argembeau, A. (2020). Event segmentation and the temporal compression of experience in episodic memory. *Psychological Research*, *84*(2), 481–490. https://doi.org/10.1007/s00426-018-1047-y

Jeunehomme, O., Folville, A., Stawarczyk, D., Van der Linden, M., & D'Argembeau, A. (2018). Temporal compression in episodic memory for real-life events. *Memory*, *26*(6), 759–770. https://doi.org/10.1080/09658211.2017.1406120

Jiang, L., Mao, R., Zhou, Q., Yang, Y., Cao, J., Ding, Y., Yang, Y., Zhang, X., Li, L., & Xu, L. (2016). Inhibition of Rac1 activity in the hippocampus impairs the forgetting of contextual fear memory. *Molecular Neurobiology*, *53*(2), 1247–1253. https://doi.org/10.1007/s12035-015-9093-6

Josselyn, S. A., Shi, C., Carlezon, W. A., Neve, R. L., Nestler, E. J., & Davis, M. (2001). Long-term memory is facilitated by cAMP response element-binding protein overexpression in the amygdala. *Journal of Neuroscience*, *21*(7), 2404–2412. https://doi.org/10.1523/JNEUROSCI.21-07-02404.2001

Kahana, M. J. (1996). Associative retrieval processes in free recall. *Memory and Cognition*, *24*(1), 103–109. https://doi.org/10.3758/BF03197276

Kahana, M. J. (2020). Computational models of memory search. *Annual Review of Psychology*, *71*, 107–138. https://doi.org/10.1146/annurev-psych-010418-103358

Kalle Kossio, Y. F., Goedeke, S., Klos, C., & Memmesheimer, R. M. (2021). Drifting assemblies for persistent memory: Neuron transitions and unsupervised compensation. *Proceedings of the National Academy of Sciences of the United States of America*, *118*(46), Article e2023832118. https://doi.org/10.1073/pnas.2023832118

Kang, L., & Toyoizumi, T. (2024). Distinguishing examples while building concepts in hippocampal and artificial networks. *Nature Communications*, *15*(1), Article 647. https://www.nature.com/articles/s41467-024-44877-0#citeas

Karlsson Wirebring, L., Wiklund-Hornqvist, C., Eriksson, J., Andersson, M., Jonsson, B., & Nyberg, L. (2015). Lesser neural pattern similarity across repeated tests is associated with better long-term memory retention. *Journal of Neuroscience*, *35*(26), 9595–9602. https://doi.org/10.1523/JNEUROSCI.3550-14.2015

Karpicke, J. D., & Bauernschmidt, A. (2011). Spaced retrieval: Absolute spacing enhances learning regardless of relative spacing. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *37*(5), 1250–1257. https://doi.org/10.1037/a0023436

Karpicke, J. D., Lehman, M., & Aue, W. R. (2014). Retrieval-based learning: An episodic context account, In B. H. Ross (Ed.), *Psychology of learning and motivation* (Vol. 61, pp. 237–284). Academic Press.

Karpicke, J. D., & Roediger, H. L. (2010). Is expanding retrieval a superior method for learning text materials? *Memory and Cognition*, *38*(1), 116–124. https://doi.org/10.3758/MC.38.1.116

Katlowitz, K. A., Picardo, M. A., & Long, M. A. (2018). Stable sequential activity underlying the maintenance of a precisely executed skilled behavior. *Neuron*, *98*(6), 1133–1140.e3. https://doi.org/10.1016/j.neuron.2018.05.017

Keinath, A. T., Mosser, C.-A., & Brandon, M. P. (2022). The representation of context in mouse hippocampus is preserved despite neural drift. *Nature Communications*, *13*, Article 2415. https://doi.org/10.1038/s41467-022-30198-7

Ketz, N., Morkonda, S. G., & O'Reilly, R. C. (2013). Theta coordinated error-driven learning in the hippocampus. *PLOS Computational Biology*, *9*(6), Article e1003067. https://doi.org/10.1371/journal.pcbi.1003067

Kiliç, A., Criss, A. H., & Howard, M. W. (2013). A causal contiguity effect that persists across time scales. *Journal of Experimental Psychology: Learning Memory and Cognition*, *39*(1), 297–303. https://doi.org/10.1037/a0028463

Kinsky, N. R., Sullivan, D. W., Mau, W., Hasselmo, M. E., & Eichenbaum, H. B. (2018). Hippocampal place fields maintain a coherent and flexible map across long timescales. *Current Biology*, *28*(22), 3578–3588.e6. https://doi.org/10.1016/j.cub.2018.09.037

Kitamura, T., Ogawa, S. K., Roy, D. S., Okuyama, T., Morrissey, M. D., Smith, L. M., Redondo, R. L., & Tonegawa, S. (2017). Engrams and circuits crucial for systems consolidation of a memory. *Science*, *356*(6333), 73–78. https://doi.org/10.1126/science.aam6808

Kording, K. P., Tenenbaum, J. B., & Shadmehr, R. (2007). The dynamics of memory as a consequence of optimal adaptation to a changing body. *Nature Neuroscience*, *10*(6), 779–786. https://doi.org/10.1038/nn1901

Kowadlo, G., Ahmed, A., & Rawlinson, D. (2019). *AHA! an 'Artificial Hippocampal Algorithm' for episodic machine learning*. arXiv. https://doi.org/10.48550/arXiv.1909.10340

Kramár, E. A., Babayan, A. H., Gavin, C. F., Cox, C. D., Jafari, M., Gall, C. M., Rumbaugh, G., & Lynch, G. (2012). Synaptic evidence for the efficacy of spaced learning. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(13), 5121–5126. https://doi.org/10.1073/pnas.1120700109

Krueger, W. (1929). The effect of overlearning on retention. *Journal of Experimental Psychology*, *12*(142), 71–78. https://doi.org/10.1037/h0072036

Krug, D., Davis, T. B., & Glover, J. A. (1990). Massed versus distributed repeated reading: A case of forgetting helping recall? *Journal of Educational Psychology*, *82*(2), 366–371. https://doi.org/10.1037/0022-0663.82.2.366

Ku, S.-P., Hargreaves, E. L., Wirth, S., & Suzuki, W. A. (2021). The contributions of entorhinal cortex and hippocampus to error driven learning. *Communications Biology*, *4*, Article 618. https://doi.org/10.1038/s42003-021-02096-z

Kumaran, D., & McClelland, J. L. (2012). Generalization through the recurrent interaction of episodic memories: A model of the hippocampal system. *Psychological Review*, *119*(3), 573–616. https://doi.org/10.1037/a0028681

Küpper-Tetzel, C. E., & Erdfelder, E. (2012). Encoding, maintenance, and retrieval processes in the lag effect: A multinomial processing tree analysis. *Memory*, *20*(1), 37–47. https://doi.org/10.1080/09658211.2011.631550

Küpper-Tetzel, C. E., Kapler, I. V., & Wiseheart, M. (2014). Contracting, equal, and expanding learning schedules: The optimal distribution of learning sessions depends on retention interval. *Memory and Cognition*, *42*(5), 729–741. https://doi.org/10.3758/s13421-014-0394-1

Landauer, T. K. (1969). Reinforcement as consolidation. *Psychological Review*, *76*(1), 82–96. https://doi.org/10.1037/h0026746

Landauer, T. K., & Bjork, R. A. (1978). Optimum rehearsal patterns and name learning. In M. Gruneberg, P. Morris, & R. Sykes, (Eds.), *Practical aspects of memory* (pp. 625–632). Academic Press.

Lee, J. S., Briguglio, J. J., Cohen, J. D., Romani, S., & Lee, A. K. (2020). The statistical structure of the hippocampal code for space as a function of time, context, and value. *Cell*, *183*(3), 620–635.e22. https://doi.org/10.1016/j.cell.2020.09.024

Lee, T. D., & Genovese, E. D. (1988). Distribution of practice in motor skill acquisition: Learning and performance effects reconsidered. *Research Quarterly for Exercise and Sport*, *59*(4), 277–287. https://doi.org/10.1080/02701367.1988.10609373

Leutgeb, S., & Leutgeb, J. K. (2007). Pattern separation, pattern completion, and new neuronal codes within a continuous CA3 map. *Learning & Memory*, *14*(11), 745–57. https://doi.org/10.1101/lm.703907

Leutgeb, S., Leutgeb, J. K., Treves, A., Moser, M. B., & Moser, E. I. (2004). Distinct ensemble codes in hippocampal areas CA3 and CA1. *Science*, *305*(5688), 1295–1298. https://doi.org/10.1126/science.1100265

Lewandowsky, S., Ecker, U. K., Farrell, S., & Brown, G. D. (2012). Models of cognition and constraints from neuroscience: A case study involving consolidation. *Australian Journal of Psychology*, *64*(1), 37–45. https://doi.org/10.1111/j.1742-9536.2011.00042.x

Lewis, P. A., & Miall, R. C. (2009). The precision of temporal judgement: Milliseconds, many minutes, and beyond. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1525), 1897–1905. https://doi.org/10.1098/rstb.2009.0020

Li, C., & Yang, J. (2020). Role of the hippocampus in the spacing effect during memory retrieval. *Hippocampus*, *30*(7), 703–714. https://doi.org/10.1002/hipo.23193

Liu, Y., Levy, S., Mau, W., Geva, N., Rubin, A., Ziv, Y., Hasselmo, M. E., & Howard, M. W. (2022). Consistent population activity on the scale of minutes in the mouse hippocampus. *Hippocampus*, *32*(5), 359–372. https://doi.org/10.1002/hipo.23409

Liu, Y., Tiganj, Z., Hasselmo, M. E., & Howard, M. W. (2019). A neural microcircuit model for a scalable scale-invariant representation of time. *Hippocampus*, *29*(3), 260–274. https://doi.org/10.1002/hipo.22994

Lohnas, L. J., Polyn, S. M., & Kahana, M. J. (2011). Contextual variability in free recall. *Journal of Memory and Language*, *64*(3), 249–255. https://doi.org/10.1016/j.jml.2010.11.003

Lohnas, L. J., Polyn, S. M., & Kahana, M. J. (2015). Expanding the scope of memory search: Modeling intralist and interlist effects in free recall. *Psychological Review*, *122*(2), 337–363. https://doi.org/10.1037/a0039036

Lositsky, O., Chen, J., Toker, D., Honey, C. J., Shvartsman, M., Poppenk, J. L., Hasson, U., & Norman, K. A. (2016). Neural pattern change during encoding of a narrative predicts retrospective duration estimates. *eLife*, *5*, Article 16070. https://doi.org/10.7554/eLife.16070

Lu, Q., Hasson, U., & Norman, K. A. (2020). A neural network model of when to retrieve and encode episodic memories. *eLife*, *11*, Article e74445. https://doi.org/10.7554/eLife.74445

MacDonald, C. J., Lepage, K. Q., Eden, U. T., & Eichenbaum, H. (2011). Hippocampal "time cells" bridge the gap in memory for discontiguous events. *Neuron*, *71*(4), 737–49. https://doi.org/10.1016/j.neuron.2011.07.012

Maddox, G. B. (2016). Understanding the underlying mechanism of the spacing effect in verbal learning: A case for encoding variability and study-phase retrieval. *Journal of Cognitive Psychology*, *28*(6), 684–706. https://doi.org/10.1080/20445911.2016.1181637

Mankin, E. A., Diehl, G. W., Sparks, F. T., Leutgeb, S., & Leutgeb, J. K. (2015). Hippocampal CA2 activity patterns change over time to a larger extent than between spatial contexts. *Neuron*, *85*(1), 190–201. https://doi.org/10.1016/j.neuron.2014.12.001

Manning, J. R., Polyn, S. M., Baltuch, G. H., Litt, B., & Kahana, M. J. (2011). Oscillatory patterns in temporal lobe reveal context reinstatement during memory search. *Proceedings of the National Academy of Sciences*, *108*(31), 12893–12897. https://doi.org/10.1073/pnas.1015174108

Manns, J. R., Howard, M. W., & Eichenbaum, H. (2007). Gradual changes in hippocampal activity support remembering the order of events. *Neuron*, *56*(3), 530–540. https://doi.org/10.1016/j.neuron.2007.08.017

Margolis, D. J., Lütcke, H., Schulz, K., Haiss, F., Weber, B., Kügler, S., Hasan, M. T., & Helmchen, F. (2012). Reorganization of cortical population activity imaged throughout long-term sensory deprivation. *Nature Neuroscience*, *15*(11), 1539–1546. https://doi.org/10.1038/nn.3240

Marks, T. D., & Goard, M. J. (2021). Stimulus-dependent representational drift in primary visual cortex. *Nature Communications*, *12*(1), Article 5169. https://doi.org/10.1038/s41467-021-25436-3

Marr, D. (1971). Simple memory: A theory for archicortex. *Philosophical Transactions of the Royal Society of London: Series B, Biological Sciences*, *262*(841), 23–81. https://doi.org/10.1098/rstb.1971.0078

Maskarinec, A. S., & Thompson, C. P. (1976). The within-list distributed practice effect: Tests of the varied context and varied encoding hypotheses. *Memory & Cognition*, *4*(6), 741–746. https://doi.org/10.3758/BF03213242

Masse, N. Y., Grant, G. D., & Freedman, D. J. (2018). Alleviating catastrophic forgetting using contextdependent gating and synaptic stabilization. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(44), E104657–E104675. https://doi.org/10.1073/pnas.1803839115

Mau, W., Hasselmo, M. E., & Cai, D. J. (2020). The brain in motion: How ensemble fluidity drives memory-updating and flexibility. *eLife*, *9*, Article e63550. https://doi.org/10.7554/eLife.63550

Mau, W., Sullivan, D. W., Kinsky, N. R., Hasselmo, M. E., Howard, M. W., & Eichenbaum, H. (2018). The same hippocampal CA1 population simultaneously codes temporal information over multiple timescales. *Current Biology*, *28*(10), 1499–1508.e4. https://doi.org/10.1016/j.cub.2018.03.051

Maxcey, A. M., Shiffrin, R. M., Cousineau, D., & Atkinson, R. C. (2022). Two case studies of very long-term retention. *Psychonomic Bulletin & Review*, *29*(2), 563–567. https://doi.org/10.3758/s13423-021-02002-y

McClelland, J. L., McNaughton, B. L., & O'Reilly, R. C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. *Psychological Review*, *102*(3), 419–57. https://doi.org/10.1037/0033-295X.102.3.419

McFarland, C. E., Rhodes, D. D., & Frey, T. J. (1979). Semantic-feature variability and the spacing effect. *Journal of Verbal Learning and Verbal Behavior*, *18*(2), 163–172. https://doi.org/10.1016/S0022-5371(79)90100-2

McGuire, K. L., Amsalem, O., Sugden, A. U., Ramesh, R. N., Fernando, J., Burgess, C. R., & Andermann, M. L. (2022). Visual association cortex links cues with conjunctions of reward and locomotor contexts. *Current Biology*, *32*(7), 1563–1576. https://doi.org/10.1016/j.cub.2022.02.028

Melton, A. W. (1970). The situation with respect to the spacing of repetitions and memory. *Journal of Verbal Learning and Verbal Behavior*, *9*(5), 596–606. https://doi.org/10.1016/S0022-5371(70)80107-4

Mensink, G.-J., & Raaijmakers, J. G. (1988). A model for interference and forgetting. *Psychological Review*, *95*(4), 434–55. https://doi.org/10.1037/0033-295X.95.4.434

Merchant, H., Harrington, D. L., & Meck, W. H. (2013). Neural basis of the perception and estimation of time. *Annual Review of Neuroscience*, *36*, 313–336. https://doi.org/10.1146/annurev-neuro-062012-170349

Metcalfe, J. (2017). Learning from errors. *Annual Review of Psychology*, *68*, 465–489. https://doi.org/10.1146/annurev-psych-010416-044022

Mettler, E., Massey, C. M., & Kellman, P. J. (2016). A comparison of adaptive and fixed schedules of practice. *Journal of Experimental Psychology: General*, *145*(7), 897–917. https://doi.org/10.1037/xge0000170

Miall, C. (1989). The storage of time intervals using oscillating neurons. *Neural Computation*, *1*(3), 359–371. https://doi.org/10.1162/neco.1989.1.3.359

Michelmann, S., Price, A. R., Aubrey, B., Doyle, W. K., Friedman, D., Dugan, P. C., Devinsky, O., Devore, S., Flinker, A., Hasson, U., & Norman, K. A. (2021). Moment-by-moment tracking of naturalistic learning and its underlying hippocampo-cortical interactions. *Nature Communications*, *12*, Article 5394. https://doi.org/10.1038/s41467-021-25376-y

Montchal, M. E., Reagh, Z. M., & Yassa, M. A. (2019). Precise temporal memories are supported by the lateral entorhinal cortex in humans. *Nature*

Neuroscience, *22*(2), 284–288. https://doi.org/10.1038/s41593-018-0303-1

Moreton, B. J., & Ward, G. (2010). Time scale similarity and long-term memory for autobiographical events. *Psychonomic Bulletin and Review*, *17*(4), 510–515. https://doi.org/10.3758/PBR.17.4.510

Mozer, M. C., Pashler, H., Cepeda, N., Lindsey, R., & Vul, E. (2009). Predicting the optimal spacing of study: A multiscale context model of memory. In Y. Bengio, D. Schuurmans, J. Lafferty, C. K. I. Williams, & A. Culotta (Eds.), *Advances in neural information processing systems* (Vol. 22, pp. 1321–1329). NIPS Foundation.

Mundorf, A. M., Lazarus, L. T., Uitvlugt, M. G., & Healey, M. K. (2021). A test of retrieved context theory: Dynamics of recall after incidental encoding. *Journal of Experimental Psychology: Learning Memory and Cognition*, *47*(8), 1264–1287. https://doi.org/10.1037/xlm0001001

Murdock, B. B. (1997). Context and mediators in a theory of distributed associative memory (TODAM2). *Psychological Review*, *104*(4), 839–862. https://doi.org/10.1037/0033-295X.104.4.839

Murray, J. M., & Escola, G. S. (2020). Remembrance of things practiced with fast and slow learning in cortical and subcortical pathways. *Nature Communications*, *11*(1), Article 6441. https://doi.org/10.1038/s41467-020-19788-5

Nadel, L., & Moscovitch, M. (1997). Memory consolidation, retrograde amnesia and the hippocampal complex. *Current Opinion in Neurobiology*, *7*(2), 217–27. https://doi.org/10.1016/S0959-4388(97)80010-4

Nairne, J. S. (2002). The myth of the encoding-retrieval match. *Memory*, *10*(5–6), 389–395. https://doi.org/10.1080/09658210244000216

Nakashiba, T., Young, J. Z., McHugh, T. J., Buhl, D. L., & Tonegawa, S. (2008). Transgenic inhibition of synaptic transmission reveals role of CA3 output in hippocampal learning. *Science*, *319*(5867), 1260–1264. https://doi.org/10.1126/science.1151120

Nielson, D. M., Smith, T. A., Sreekumar, V., Dennis, S., & Sederberg, P. B. (2015). Human hippocampus represents space and time during retrieval of real-world memories. *Proceedings of the National Academy of Sciences*, *112*(35), 11078–83. https://doi.org/10.1073/pnas.1507104112

Niv, Y., Daniel, R., Geana, A., Gershman, S. J., Leong, Y. C., Radulescu, A., & Wilson, R. C. (2015). Reinforcement learning in multidimensional environments relies on attention mechanisms. *The Journal of Neuroscience*, *35*(21), 8145–8157. https://doi.org/10.1523/JNEUROSCI.2978-14.2015

Nonaka, M., Fitzpatrick, R., Lapira, J., Wheeler, D., Spooner, P. A., Corcoles-Parada, M., Muñoz-López, M., Tully, T., Peters, M., & Morris, R. G. (2017). Everyday memory: Towards a translationally effective method of modelling the encoding, forgetting and enhancement of memory. *European Journal of Neuroscience*, *46*(4), 1937–1953. https://doi.org/10.1111/ejn.13637

Norman, K. A., & O'Reilly, R. C. (2003). Modeling hippocampal and neocortical contributions to recognition memory: A complementary-learning-systems approach. *Psychological Review*, *110*(4), 611–46. https://doi.org/10.1037/0033-295X.110.4.611

Norman, Y., Raccah, O., Liu, S., Parvizi, J., & Malach, R. (2021). Hippocampal ripples and their coordinated dialogue with the default mode network during recent and remote recollection. *Neuron*, *109*(17), 2767–2780.e5. https://doi.org/10.1016/j.neuron.2021.06.020

Noulhiane, M., Pouthas, V., Hasboun, D., Baulac, M., & Samson, S. (2007). Role of the medial temporal lobe in time estimation in the range of minutes. *NeuroReport*, *18*(10), 1035–1038. https://doi.org/10.1097/WNR.0b013e3281668be1

O'Reilly, R. C. (1996). Biologically plausible error-driven learning using local activation differences: The generalized recirculation algorithm. *Neural Computation*, *8*(5), 895–938. https://doi.org/10.1162/neco.1996.8.5.895

O'Reilly, R. C., & McClelland, J. L. (1994). Hippocampal conjunctive encoding, storage, and recall: Avoiding a trade-off. *Hippocampus*, *4*(6), 661–682. https://doi.org/10.1002/hipo.450040605

O'Reilly, R. C., & Munakata, Y. (2000). *Computational explorations in cognitive neuroscience: Understanding the mind by simulating the brain*. MIT Press.

O'Reilly, R. C., & Rudy, J. W. (2001). Conjunctive representations in learning and memory: Principles of cortical and hippocampal function. *Psychological Review*, *108*(2), 311–345. https://doi.org/10.1037/0033-295X.108.2.311

O'Reilly, R. C., Wyatte, D. R., & Rohrlich, J. (2017). *Deep predictive learning: A comprehensive model of three visual streams*. arXiv. https://arxiv.org/abs/1709.04654

Osth, A. F., & Fox, J. (2019). Are associations formed across pairs? A test of learning by temporal contiguity in associative recognition. *Psychonomic Bulletin and Review*, *26*(5), 1650–1656. https://doi.org/10.3758/s13423-019-01616-7

Parker, A., Dagnall, N., & Coyle, A. M. (2007). Environmental context effects in conceptual explicit and implicit memory. *Memory*, *15*(4), 423–434. https://doi.org/10.1080/09658210701309834

Pastalkova, E., Itskov, V., Amarasingham, A., & Buzsáki, G. (2008). Internally generated cell assembly sequences in the rat hippocampus. *Science*, *321*(5894), 1322–1377. https://doi.org/10.1126/science.1159775

Pereira, A. C., Huddleston, D. E., Brickman, A. M., Sosunov, A. A., Hen, R., McKhann, G. M., Sloan, R., Gage, F. H., Brown, T. R., & Small, S. A. (2007). An in vivo correlate of exercise-induced neurogenesis in the adult dentate gyrus. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(13), 5638–5643. https://doi.org/10.1073/pnas.0611721104

Pérez-Ortega, J., Alejandre-García, T., & Yuste, R. (2021). Long-term stability of cortical ensembles. *elife*, *10*, Article e64449. https://doi.org/10.7554/eLife.64449

Peterson, L. R., Wampler, R., Kirkpatrick, M., & Saltzman, D. (1963). Effect of spacing presentations on retention of a paired associate over short intervals. *Journal of Experimental Psychology*, *66*(2), 206–209. https://doi.org/10.1037/h0046694

Polyn, S. M., Norman, K. A., & Kahana, M. J. (2009). A context maintenance and retrieval model of organization. *Psychological Review*, *116*(1), 129–156. https://doi.org/10.1037/a0014420

Pyc, M. A., & Rawson, K. A. (2009). Testing the retrieval effort hypothesis: Does greater difficulty correctly recalling information lead to higher levels of memory? *Journal of Memory and Language*, *60*(4), 437–47. https://doi.org/10.1016/j.jml.2009.01.004

Quiroga, R. Q., Reddy, L., Kreiman, G., Koch, C., & Fried, I. (2005). Invariant visual representation by single neurons in the human brain. *Nature*, *435*(7045), 1102–1107. https://doi.org/10.1038/nature03687

Raaijmakers, J. G. (2003). Spacing and repetition effects in human memory: Application of the SAM model. *Cognitive Science*, *27*(3), 431–452. https://doi.org/10.1207/s15516709cog2703_5

Radvansky, G. A., Krawietz, S. A., & Tamplin, A. K. (2011). Walking through doorways causes forgetting: Further explorations. *Quarterly Journal of Experimental Psychology*, *64*(8), 1632–1645. https://doi.org/10.1080/17470218.2011.571267

Rajji, T., Chapman, D., Eichenbaum, H., & Greene, R. (2006). The role of CA3 hippocampal NMDA receptors in paired associate learning. *Journal of Neuroscience*, *26*(3), 908–915. https://doi.org/10.1523/JNEUROSCI.4194-05.2006

Rashid, A. J., Yan, C., Mercaldo, V., Hsiang, H. L., Park, S., Cole, C. J., De Cristofaro, A., Yu, J., Ramakrishnan, C., Lee, S. Y., Deisseroth, K., Frankland, P. W., & Josselyn, S. A. (2016). Competition between engrams influences fear memory formation and recall. *Science*, *353*(6297), 383–387. https://doi.org/10.1126/science.aaf0594

Rawson, K. A., & Dunlosky, J. (2011). Optimizing schedules of retrieval practice for durable and efficient learning: How much is enough? *Journal of Experimental Psychology: General*, *140*(3), 283–302. https://doi.org/10.1037/a0023956

Rawson, K. A., & Kintsch, W. (2005). Rereading effects depend on time of test. *Journal of Educational Psychology*, *97*(1), 70–80. https://doi.org/10.1037/0022-0663.97.1.70

Rawson, K. A., Vaughn, K. E., Walsh, M., & Dunlosky, J. (2018). Investigating and explaining the effects of successive relearning on long-term retention. *Journal of Experimental Psychology: Applied*, *24*(1), 57–71. https://doi.org/10.1037/xap0000146

Reagh, Z. M., Noche, J. A., Tustison, N. J., Delisle, D., Murray, E. A., & Yassa, M. A. (2018). Functional imbalance of anterolateral entorhinal cortex and hippocampal dentate/CA3 underlies age-related object pattern separation deficits. *Neuron*, *97*(5), 1187–1198. https://doi.org/10.1016/j.neuron.2018.01.039

Reddy, L., Zoefel, B., Possel, J. K., Peters, J., Dijksterhuis, D. E., Poncet, M., van Straaten, E. C., Baayen, J. C., Idema, S., & Self, M. W. (2021). Human hippocampal neurons track moments in a sequence of events. *Journal of Neuroscience*, *41*(31), 6714–6725. https://doi.org/10.1523/JNEUROSCI.3157-20.2021

Remme, M. W., Bergmann, U., Alevi, D., Schreiber, S., Sprekeler, H., & Kempter, R. (2021). Hebbian plasticity in parallel synaptic pathways: A circuit mechanism for systems memory consolidation. *PLOS Computational Biology*, *17*(12), Article e1009681. https://doi.org/10.1371/journal.pcbi.1009681

Rescorla, R., & Wagner, A. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A. Black & W. Prokasy (Eds.), *Classical conditioning II: Current research and theory* (pp. 64–99). Appleton-Century-Crofts.

Rogerson, T., Cai, D. J., Frank, A., Sano, Y., Shobe, J., Lopez-Aranda, M. F., & Silva, A. J. (2014). Synaptic tagging during memory allocation. *Nature Reviews Neuroscience*, *15*(3), 157–69. https://doi.org/10.1038/nrn3667

Rohrer, D. (2009). The effects of spacing and mixing practice problems. *Journal for Research in Mathematics Education*, *40*(1), 4–17. https://doi.org/10.5951/jresematheduc.40.1.0004

Rolls, E. T. (1989). Functions of neuronal networks in the hippocampus and neocortex in memory. *Neural models of plasticity* (pp. 240–265). Elsevier.

Rolls, E. T., & Kesner, R. P. (2006). A computational theory of hippocampal function, and empirical tests of the theory. *Progress in Neurobiology*, *79*(1), 1–48. https://doi.org/10.1016/j.pneurobio.2006.04.005

Rolls, E. T., & Mills, P. (2019). The generation of time in the hippocampal memory system. *Cell Reports*, *28*(7), 1649–1658. https://doi.org/10.1016/j.celrep.2019.07.042

Ross, B. H., & Landauer, T. K. (1978). Memory for at least one of two items: Test and failure of several theories of spacing effects. *Journal of Verbal Learning and Verbal Behavior*, *17*(6), 669–680. https://doi.org/10.1016/S0022-5371(78)90403-6

Rossi-Pool, R., Zizumbo, J., Alvarez, M., Vergara, J., Zainos, A., & Romo, R. (2019). Temporal signals underlying a cognitive process in the dorsal premotor cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(15), 7523–7532. https://doi.org/10.1073/pnas.1820474116

Rouhani, N., Norman, K. A., Niv, Y., & Bornstein, A. M. (2020). Reward prediction errors create event boundaries in memory. *Cognition*, *203*, Article 104269. https://doi.org/10.1016/j.cognition.2020.104269

Rowe, M. K., & Craske, M. G. (1998). Effects of an expanding-spaced vs massed exposure schedule on fear reduction and return of fear. *Behaviour Research and Therapy*, *36*(7–8), 701–717. https://doi.org/10.1016/S0005-7967(97)10016-X

Rowland, C. A. (2014). The effect of testing versus restudy on retention: A meta-analytic review of the testing effect. *Psychological Bulletin*, *140*(6), 1432–1463. https://doi.org/10.1037/a0037559

Rubin, A., Geva, N., Sheintuch, L., & Ziv, Y. (2015). Hippocampal ensemble dynamics timestamp events in long-term memory. *eLife*, *4*, Article e12247. https://doi.org/10.7554/eLife.12247

Rubin, D. C., & Wenzel, A. E. (1996). One hundred years of forgetting: A quantitative description of retention. *Psychological Review*, *103*(4), 734–760. https://doi.org/10.1037/0033-295X.103.4.734

Rudy, J. W., & O'Reilly, R. C. (2001). Conjunctive representations, the hippocampus, and contextual fear conditioning. *Cognitive, Affective, & Behavioral Neuroscience*, *1*(1), 66–82. https://doi.org/10.3758/CABN.1.1.66

Rule, M. E., & O'Leary, T. (2022). Self-healing codes: How stable neural populations can track continually reconfiguring neural representations. *Proceedings of the National Academy of Sciences of the United States of America*, *119*(7), Article e2106692119. https://doi.org/10.1073/pnas.2106692119

Rule, M. E., O'Leary, T., & Harvey, C. D. (2019). Causes and consequences of representational drift. *Current Opinion in Neurobiology*, *58*, 141–147. https://doi.org/10.1016/j.conb.2019.08.005

Russo, R., Parkin, A. J., Taylor, S. R., & Wiks, J. (1998). Revising current two-process accounts of spacing effects in memory. *Journal of Experimental Psychology: Learning Memory and Cognition*, *24*(1), 161–172. https://doi.org/10.1037/0278-7393.24.1.161

Schapiro, A. C., Turk-Browne, N. B., Botvinick, M. M., & Norman, K. A. (2017). Complementary learning systems within the hippocampus: A neural network modelling approach to reconciling episodic memory with statistical learning. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *372*(1711), Article 20160049. https://doi.org/10.1098/rstb.2016.0049

Scharf, M. T., Woo, N. H., Matthew Lattal, K., Young, J. Z., Nguyen, P. V., & Abel, T. (2002). Protein synthesis is required for the enhancement of long-term potentiation and long-term memory by spaced training. *Journal of Neurophysiology*, *87*(6), 2770–2777. https://doi.org/10.1152/jn.2002.87.6.2770

Sederberg, P. B., Gershman, S. J., Polyn, S. M., & Norman, K. A. (2011). Human memory reconsolidation can be explained using the temporal context model. *Psychonomic Bulletin & Review*, *18*(3), 455–468. https://doi.org/10.3758/s13423-011-0086-9

Sederberg, P. B., Howard, M. W., & Kahana, M. J. (2008). A context-based theory of recency and contiguity in free recall. *Psychological Review*, *115*(4), 893–912. https://doi.org/10.1037/a0013396

Sellevoll, M., Barak, O., Morris, G., Derdikman, D., Khatib, D., Ratzon, A., Sellevoll, M., Barak, O., Morris, G., & Derdikman, D. (2023). Active experience, not time, determines within-day representational drift in dorsal CA1. *Neuron*, *111*(5), 2348–2356.e5. https://doi.org/10.1016/j.neuron.2023.05.014

Shea, C. H., Lai, Q., Black, C., & Park, J. H. (2000). Spacing practice sessions across days benefits the learning of motor skills. *Human Movement Science*, *19*(5), 737–760. https://doi.org/10.1016/S0167-9457(00)00021-X

Shen, Y., Zhou, M., Cai, D., Filho, D. A., Fernandes, G., Cai, Y., Kim, N., Necula, D., Zhou, C., Liu, A., Kang, X., Kamata, M., Lavi, A., Huang, S., Silva, T., Heo, W. D., & Silva, A. J. (2022). CCR5 closes the temporal window for memory linking. *Nature*, *606*(7912), 146–152 https://doi.org/10.1038/s41586-022-04783-1

Shimbo, A., Izawa, E. I., & Fujisawa, S. (2021). Scalable representation of time in the hippocampus. *Science Advances*, *7*(6), Article eabd7013. https://doi.org/10.1126/sciadv.abd7013

Siegle, J. H., & Wilson, M. A. (2014). Enhancement of encoding and retrieval functions through theta phase-specific manipulation of hippocampus. *eLife*, *3*, Article e03061 https://doi.org/10.7554/eLife.03061

Sietsma, J., & Dow, R. J. (1991). Creating artificial neural networks that generalize. *Neural Networks*, *4*(1), 67–79. https://doi.org/10.1016/0893-6080(91)90033-2

Singer, A. C., Karlsson, M. P., Nathe, A. R., Carr, M. F., & Frank, L. M. (2010). Experience-dependent development of coordinated hippocampal spatial activity representing the similarity of related locations.

Journal of Neuroscience, *30*(35), 11586–11604. https://doi.org/10.1523/JNEUROSCI.0926-10.2010

Singh, I., Oliva, A., & Howard, M. W. (2017). *Visual memories are stored along a logarithmically-compressed representation of the past*. bioRxiv. https://doi.org/10.1101/101295

Skinner, B. (1938). *The behavior of organisms*. Appleton-Century-Crofts.

Smith, C. D., & Scarf, D. (2017). Spacing repetitions over long timescales: A review and a reconsolidation explanation. *Frontiers in Psychology*, *8*, Article 962. https://doi.org/10.3389/fpsyg.2017.00962

Smith, S. M. (1982). Enhancement of recall using multiple environmental contexts during learning. *Memory & Cognition*, *10*(5), 405–412. https://doi.org/10.3758/BF03197642

Smith, S. M., Glenberg, A., & Bjork, R. A. (1978). Environmental context and human memory. *Memory & Cognition*, *6*(4), 342–353. https://doi.org/10.3758/BF03197465

Smith, S. M., & Handy, J. D. (2014). Effects of varied and constant environmental contexts on acquisition and retention. *Journal of Experimental Psychology: Learning Memory and Cognition*, *40*(6), 1582–1593. https://doi.org/10.1037/xlm0000019

Smith, S. M., & Handy, J. D. (2016). The crutch of context-dependency: Effects of contextual support and constancy on acquisition and retention. *Memory*, *24*(8), 1134–1141. https://doi.org/10.1080/09658211.2015.1071852

Smith, S. M., & Rothkopf, E. Z. (1984). Contextual enrichment and distribution of practice in the classroom. *Cognition and Instruction*, *1*(3), 341–358. https://doi.org/10.1207/s1532690xci0103_4

Smolen, P., Zhang, Y., & Byrne, J. H. (2016). The right time to learn: Mechanisms and optimization of spaced learning. *Nature Reviews Neuroscience*, *17*(2), 77–88. https://doi.org/10.1038/nrn.2015.18

Soderstrom, N. C., & Bjork, R. A. (2015). Learning versus performance: An integrative review. *Perspectives on Psychological Science*, *10*(2), 176–199. https://doi.org/10.1177/1745691615569000

Speer, M. E., Ibrahim, S., Schiller, D., & Delgado, M. R. (2021). Finding positive meaning in memories of negative events adaptively updates memory. *Nature Communications*, *12*(1), Article 6601. https://doi.org/10.1038/s41467-021-26906-4

Spieler, D. H., & Balota, D. A. (1996). Characteristics of associative learning in younger and older adults: Evidence from an episodic priming paradigm. *Psychology and Aging*, *11*(4), 607–620. https://doi.org/10.1037/0882-7974.11.4.607

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, *15*, 1929–1958.

Stensola, H., Stensola, T., Solstad, T., Frøland, K., Moser, M. B., & Moser, E. I. (2012). The entorhinal grid map is discretized. *Nature*, *492*(7427), 72–78. https://doi.org/10.1038/nature11649

Sun, C., Yang, W., Martin, J., & Tonegawa, S. (2020). Hippocampal neurons represent events as transferable units of experience. *Nature Neuroscience*, *23*(5), 651–663. https://doi.org/10.1038/s41593-020-0614-x

Sutton, R. S., & Barto, A. G. (1981). Toward a modern theory of adaptive networks: Expectation and prediction. *Psychological Review*, *88*(2), 135–170. https://doi.org/10.1037/0033-295X.88.2.135

Sutton, R. S., & Barto, A. G. (1998). *Introduction to reinforcement learning* (Vol. 2). MIT press.

Tahvildari, B., Fransén, E., Alonso, A. A., & Hasselmo, M. E. (2007). Switching between "on" and "off" states of persistent activity in lateral entorhinal layer III neurons. *Hippocampus*, *17*, 257–263. https://doi.org/10.1002/hipo.20270

Talmi, D., Lohnas, L. J., & Daw, N. D. (2019). A retrieved context model of the emotional modulation of memory. *Psychological Review*, *126*(4), 455–485. https://doi.org/10.1037/rev0000132

Thios, S. J., & D'Agostino, P. R. (1976). Effects of repetition as a function of study-phase retrieval. *Journal of Verbal Learning and Verbal Behavior*, *15*(5), 529–536. https://doi.org/10.1016/0022-5371(76)90047-5

Tiganj, Z., Hasselmo, M. E., & Howard, M. W. (2015). A simple biophysically plausible model for long time constants in single neurons. *Hippocampus*, *25*(1), 27–37. https://doi.org/10.1002/hipo.22347

Toppino, T. C., & Gerbier, E. (2014). *About practice: Repetition, spacing, and abstraction* (Vol. 60). Elsevier.

Toppino, T. C., Phelan, H. A., & Gerbier, E. (2018). Level of initial training moderates the effects of distributing practice over multiple days with expanding, contracting, and uniform schedules: Evidence for study-phase retrieval. *Memory and Cognition*, *46*(6), 969–978. https://doi.org/10.3758/s13421-018-0815-7

Tran, L. M., Santoro, A., Liu, L., Josselyn, S. A., Richards, B. A., & Frankland, P. W. (2022). Adult neurogenesis acts as a neural regularizer. *Proceedings of the National Academy of Sciences*, *119*(45), Article e2206704119. https://doi.org/10.1073/pnas.2206704119

Trask, S., & Bouton, M. E. (2018). Retrieval practice after multiple context changes, but not long retention intervals, reduces the impact of a final context change on instrumental behavior. *Learning and Behavior*, *46*(2), 213–221. https://doi.org/10.3758/s13420-017-0304-z

Treves, A., & Rolls, E. T. (1994). Computational analysis of the role of the hippocampus in memory. *Hippocampus*, *4*(3), 374–91. https://doi.org/10.1002/hipo.450040319

Tsao, A., Sugar, J., Lu, L., Wang, C., Knierim, J. J., Moser, M. B., & Moser, E. I. (2018). Integrating time from experience in the lateral entorhinal cortex. *Nature*, *561*(7721), 57–62. https://doi.org/10.1038/s41586-018-0459-6

Tulving, E. (1972). Episodic and semantic memory. In E. Tulving & W. Donaldson (Eds.), *Organization of memory* (pp. 381–403). Academic Press.

Uitvlugt, M. G., & Healey, M. K. (2019). Temporal proximity links unrelated news events in memory. *Psychological Science*, *30*(1), 92–104. https://doi.org/10.1177/0956797618808474

Umbach, G., Kantak, P., Jacobs, J., Kahana, M., Pfeiffer, B. E., Sperling, M., & Lega, B. (2020). Time cells in the human hippocampus and entorhinal cortex support episodic memory. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(45), 28463–28474. https://doi.org/10.1073/pnas.2013250117

Umbach, G., Tan, R., Jacobs, J., Pfeiffer, B. E., & Lega, B. (2022). Flexibility of functional neuronal assemblies supports human memory. *Nature Communications*, *13*(1), Article 6162. https://doi.org/10.1038/s41467-022-33587-0

Unsworth, N. (2008). Exploring the retrieval dynamics of delayed and final free recall: Further evidence for temporal-contextual search. *Journal of Memory and Language*, *59*(2), 223–236. https://doi.org/10.1016/j.jml.2008.04.002

Vaidya, S. P., Chitwood, R. A., & Magee, J. C. (2023). *The formation of an expanding memory representation in the hippocampus*. bioRxiv. https://doi.org/10.1101/2023.02.01.526663

van de Ven, G. M., Siegelmann, H. T., & Tolias, A. S. (2020). Brain-inspired replay for continual learning with artificial neural networks. *Nature Communications*, *11*(1), Article 4069. https://doi.org/10.1038/s41467-020-17866-2

Van den Broek, G. S. E., Takashima, A., Segers, E., Fernández, G., & Verhoeven, L. (2013). Neural correlates of testing effects in vocabulary learning. *NeuroImage*, *78*, 94–102. https://doi.org/10.1016/j.neuroimage.2013.03.071

Vazdarjanova, A., & Guzowski, J. F. (2004). Differences in hippocampal neuronal population responses to modifications of an environmental context: Evidence for distinct, yet complementary, functions of CA3 and CA1 ensembles. *The Journal of Neuroscience*, *24*(29), 6489–6496. https://doi.org/10.1523/JNEUROSCI.0350-04.2004

Verfaellie, M., Rajaram, S., Fossum, K., & Williams, L. (2008). Not all repetition is alike: Different benefits of repetition in amnesia and normal memory. *Journal of the International Neuropsychological Society*, *14*(03), 365–372. https://doi.org/10.1017/S1355617708080612

Vyleta, N. P., Borges-Merjane, C., & Jonas, P. (2016). Plasticity-dependent, full detonation at hippocampal mossy fiber-CA3 pyramidal neuron synapses. *eLife*, *5*, Article e17977. https://doi.org/10.7554/eLife.17977

Walsh, M. M., Gluck, K. A., Gunzelmann, G., Jastrzembski, T., Krusmark, M., Myung, J. I., Pitt, M. A., & Zhou, R. (2018). Mechanisms underlying the spacing effect in learning: A comparison of three computational models. *Journal of Experimental Psychology: General*, *147*(9), 1325–1348. https://doi.org/10.1037/xge0000416

Wei, X. X., Prentice, J., & Balasubramanian, V. (2015). A principle of economy predicts the functional architecture of grid cells. *eLife*, *4*, Article e08362. https://doi.org/10.7554/eLife.08362

Wen, T., & Egner, T. (2022). Retrieval context determines whether event boundaries impair or enhance temporal order memory. *Cognition*, *225*, Article 105145. https://doi.org/10.1016/j.cognition.2022.105145

Whittington, J. C., Muller, T. H., Mark, S., Chen, G., Barry, C., Burgess, N., & Behrens, T. E. (2020). The Tolman-Eichenbaum machine: Unifying space and relational memory through generalization in the hippocampal formation. *Cell*, *183*(5), 1249–1263.e23. https://doi.org/10.1016/j.cell.2020.10.024

Wigström, H. (1973). A neuron model with learning capability and its relation to mechanisms of association. *Kybernetik*, *12*(4), 204–215. https://doi.org/10.1007/BF00270573

Wiltgen, B. J., & Silva, A. J. (2007). Memory for context becomes less specific with time. *Learning & Memory*, *14*, 313–317. https://doi.org/10.1101/lm.430907

Winocur, G., Moscovitch, M., & Bontempi, B. (2010). Memory formation and long-term retention in humans and animals: Convergence towards a transformation account of hippocampal-neocortical interactions. *Neuropsychologia*, *48*(8), 2339–2356. https://doi.org/10.1016/j.neuropsychologia.2010.04.016

Witter, M. P., Doan, T. P., Jacobsen, B., Nilssen, E. S., & Ohara, S. (2017). Architecture of the entorhinal cortex a review of entorhinal anatomy in rodents with some comparative notes. *Frontiers in Systems Neuroscience*, *11*, Article 46. https://doi.org/10.3389/fnsys.2017.00046

Xue, G., Dong, Q., Chen, C., Lu, Z., Mumford, J. A., & Poldrack, R. A. (2010). Greater neural pattern similarity across repetitions is associated with better memory. *Science*, *330*(6000), 97–101. https://doi.org/10.1126/science.1193125

Xue, G., Mei, L., Chen, C., Lu, Z.-L., Poldrack, R., & Dong, Q. (2011). Spaced learning enhances subsequent recognition memory by reducing neural repetition suppression. *Journal of Cognitive Neuroscience*, *23*(7), 1624–33. https://doi.org/10.1162/jocn.2010.21532

Yaffe, R. B., Kerr, M. S. D., Damera, S., Sarma, S. V., Inati, S. K., & Zaghloul, K. A. (2014). Reinstatement of distributed cortical oscillations occurs with precise spatiotemporal dynamics during successful memory retrieval. *Proceedings of the National Academy of Sciences*, *111*(52), 18727–18732. https://doi.org/10.1073/pnas.1417017112

Yassa, M. A., & Reagh, Z. M. (2013). Competitive trace theory: A role for the hippocampus in contextual interference during retrieval. *Frontiers in Behavioral Neuroscience*, *7*, Article 107. https://doi.org/10.3389/fnbeh.2013.00107

Yeckel, M. F., & Berger, T. W. (1990). Feedforward excitation of the hippocampus by afferents from the entorhinal cortex: Redefinition of the role of the trisynaptic pathway. *Proceedings of the National Academy of Sciences of the United States of America*, *87*(15), 5832–5836. https://doi.org/10.1073/pnas.87.15.5832

Yonelinas, A. P. (2002). The nature of recollection and familiarity: A review of 30 years of research. *Journal of Memory and Language*, *46*(3), 441–517. https://doi.org/10.1006/jmla.2002.2864

Yoo, H. B., Umbach, G., & Lega, B. (2022). Episodic boundary cells in human medial temporal lobe during the free recall task. *Hippocampus*, *32*(7), 481–487. https://doi.org/10.1002/hipo.23421

Zacks, J. M., Speer, N. K., Swallow, K. M., Braver, T. S., & Reynolds, J. R. (2007). Event perception: A mind/brain perspective. *Psychological Bulletin*, *133*(2), 273–293. https://doi.org/10.1037/0033-2909.133.2.273

Zawadzka, K., Baloro, S., Wells, J., Wilding, E. L., & Hanczakowski, M. (2021). On the memory benefits of repeated study with variable tasks. *Journal of Experimental Psychology: Learning Memory and Cognition*, *47*(7), 1067–1082. https://doi.org/10.1037/xlm0001013

Zhang, K., Chen, L., Li, Y., Paez, A. G., Miao, X., Cao, D., Gu, C., Pekar, J. J., van Zijl, P. C. M., Hua, J., & Bakker, A. (2023). Differential laminar activation dissociates encoding and retrieval in the human medial and lateral entorhinal cortex. *The Journal of Neuroscience*, *43*(16), 2874–2884. https://doi.org/10.1523/JNEUROSCI.1488-22.2023

Zheng, Y., Liu, X. L., Nishiyama, S., Ranganath, C., & O'Reilly, R. C. (2022). Correcting the Hebbian mistake: Toward a fully error-driven hippocampus. *PLOS Computational Biology*, *18*(10), Article e1010589. https://doi.org/10.1371/journal.pcbi.1010589

Zhou, Y., Won, J., Karlsson, M. G., Zhou, M., Rogerson, T., Balaji, J., Neve, R., Poirazi, P., & Silva, A. J. (2009). CREB regulates excitability and the allocation of memory to subsets of neurons in the amygdala. *Nature Neuroscience*, *12*(11), 1438–1443. https://doi.org/10.1038/nn.2405

Ziv, N. E., & Brenner, N. (2018). Synaptic tenacity or lack thereof: Spontaneous remodeling of synapses. *Trends in Neurosciences*, *41*(2), 89–99. https://doi.org/10.1016/j.tins.2017.12.003

Ziv, Y., Burns, L. D., Cocker, E. D., Hamel, E. O., Ghosh, K. K., Kitch, L. J., Gamal, A. E., & Schnitzer, M. J. (2013). Long-term dynamics of CA1 hippocampal place codes. *Nature Neuroscience*, *16*(3), 264–266. https://doi.org/10.1038/nn.3329

Zou, F., Wanjia, G., Allen, E. J., Wu, Y., Charest, I., Naselaris, T., Kay, K., Kuhl, B. A., Hutchinson, J. B., & DuBrow, S. (2023). Re-expression of CA1 and entorhinal activity patterns preserves temporal context memory at long timescales. *Nature Communications*, *14*(1), Article 4350. https://doi.org/10.1038/s41467-023-40100-8

(*Appendix follows*)

Appendix

Overview

Here, we provide many important parameters for understanding the HipSTeR model architecture and critical aspects of our inputs (Tables A1, A2 and A3). See table captions for detailed definitions and descriptions. Please note that the model is available at https://github.com/JamesWardAntony/HipSTeR, with the full Leabra model and documentation available at https://github.com/CCNLab.

Network Size Parameters

Table A1
Parameters for Network Sizes

Network layer	Neuron
Input pool size	7 × 7
Input number of pools	2 × 8
ECin pool size	7 × 7
ECin number of pools	2 × 8
ECout pool size	7 × 7
ECout number of pools	2 × 8
DG size	60 × 60
CA3 size	40 × 40
CA1 pool size	20 × 20
CA1 number of pool	2 × 8

Note. The numbers for pool sizes indicate the number of neurons in each specific pool. ECin = entorhinal cortex input layer; ECout = entorhinal cortex output layer; DG = dentate gyrus; CA3 = cornu ammonis 3; CA1 = cornu ammonis 1.

Training/Testing Input Diagram

Table A2
Training/Testing Pools

TC7/OC1	TC8/OC1
TC5	TC6
TC3	TC4
TC1	TC2
B3/empty	B4/empty
B1/empty	B2/empty
A3	A4
A1	A2

Note. Each cell represents a pool in the input layer. All spacing effect simulations used TC7/TC8, whereas the decontextualization experiment used OC1/OC2. All training used B1–4, and all testing used empty pools instead. TC = temporal context; OC = other context for decontextualization experiments; B = target; A = cue; empty = no input.

(*Appendix continues*)

Parameter Changes From Default Model

Table A3

Model Parameters That Diverged From the Default (Zheng et al., 2022)

Area	Param	Value	Default
ECout	ECoutToECin.WtScale.Rel	0.4	0.5
DG	Inhib.Layer.Gi	2.95	3.8

Note. These changes were made to account for different network sizes, specifically the change from 2 × 3 input pools used in Zheng et al. (2022) to the 2 × 8 pools here. ECout = entorhinal cortex output layer; DG = dentate gyrus.

Received January 5, 2023
Revision received December 21, 2023
Accepted March 14, 2024 ■